Lecture 1: Introduction

Nayel Bettache

Department of Statistical Science, Cornell University

Machine Learning (ML) vs Stat

"Artificial intelligence (or machine learning) is actually statistics, but it uses a very gorgeous rhetoric. It is actually statistics. A lot of formulas are very old. But all artificial intelligence uses statistics to solve problems."

-Thomas J. Sargent, Nobel Prize winner in Economics, 2018

Machine Learning (ML) vs Stat

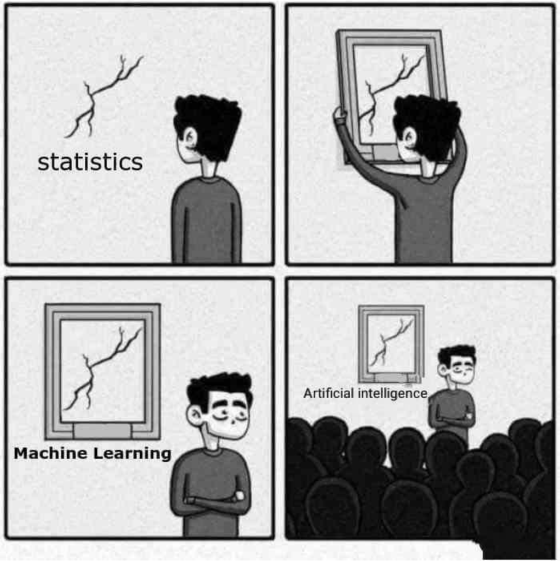
"Artificial intelligence (or machine learning) is actually statistics, but it uses a very gorgeous rhetoric. It is actually statistics. A lot of formulas are very old. But all artificial intelligence uses statistics to solve problems."

-Thomas J. Sargent, Nobel Prize winner in Economics, 2018

"When you are fundraising, it is AI. When you are hiring, it is ML. When you are implementing, it is logistic regression."

-everyone on Twitter ever

Machine Learning (ML) vs Stat



Comic copyright Thomas Wiecki, Ph.D., ODSC Europe 2018

An Overview of Statistical Learning

- Statistical learning involves tools for understanding data.

An Overview of Statistical Learning

- Statistical learning involves tools for understanding data.
- Methods are categorized as:

An Overview of Statistical Learning

- Statistical learning involves tools for understanding data.
- Methods are categorized as:
  - **Supervised learning**: Builds models to predict or estimate an output based on input data.

An Overview of Statistical Learning

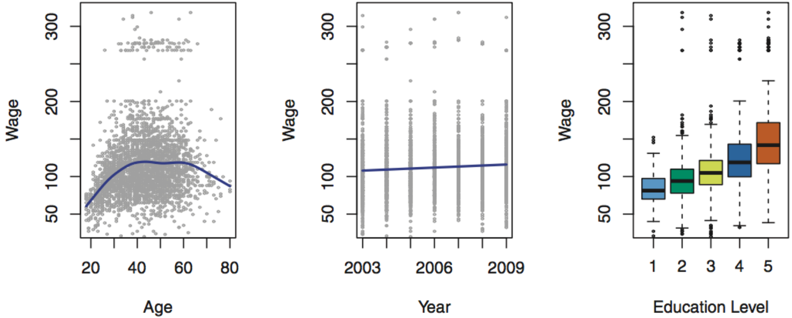
- Statistical learning involves tools for understanding data.
- Methods are categorized as:
  - **Supervised learning**: Builds models to predict or estimate an output based on input data.
  - **Unsupervised learning**: Discovers relationships and structures in data without an output variable.

An Overview of Statistical Learning

- Statistical learning involves tools for understanding data.
- Methods are categorized as:
  - **Supervised learning**: Builds models to predict or estimate an output based on input data.
  - **Unsupervised learning**: Discovers relationships and structures in data without an output variable.
- Applications span across diverse fields such as business, medicine, astrophysics, and public policy.

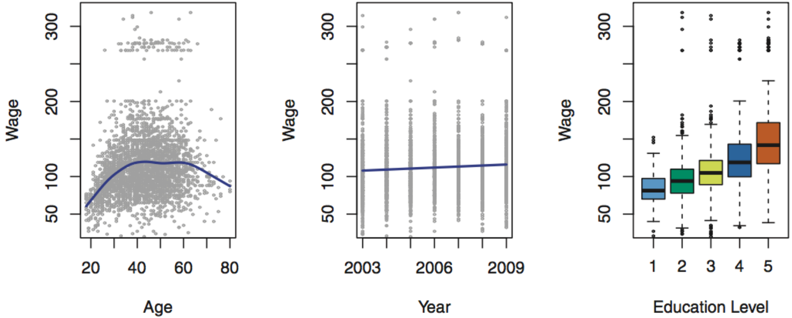
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.



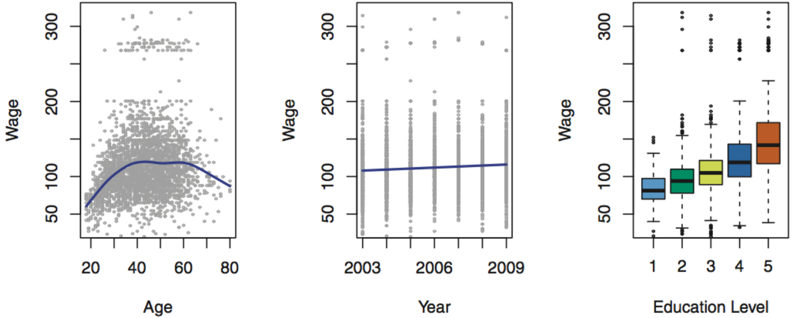
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Focus on relationships between:



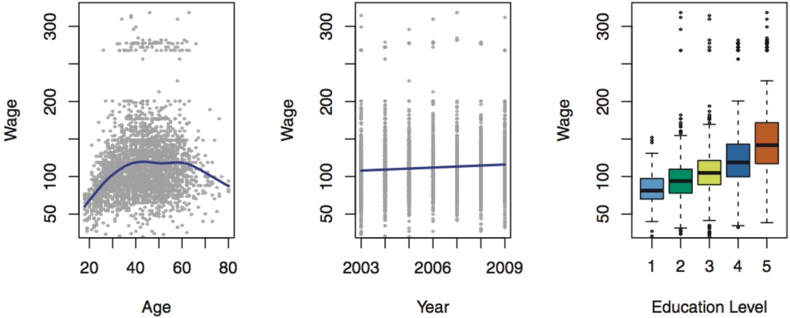
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Focus on relationships between:
  - Age and wage: Wages increase with age, then decline after age 60.



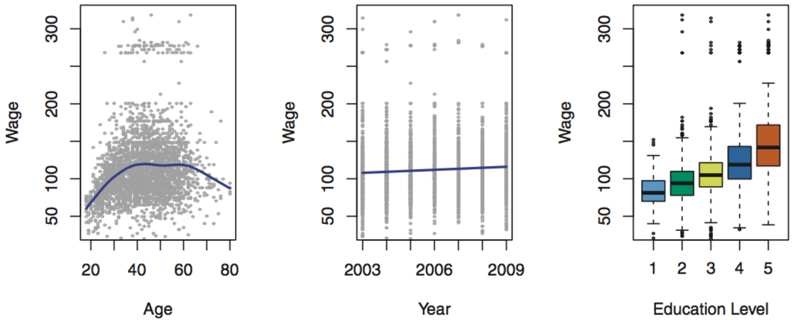
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Focus on relationships between:
    - Age and wage: Wages increase with age, then decline after age 60.
    - Year and wage: Slight wage increase between 2003 and 2009.



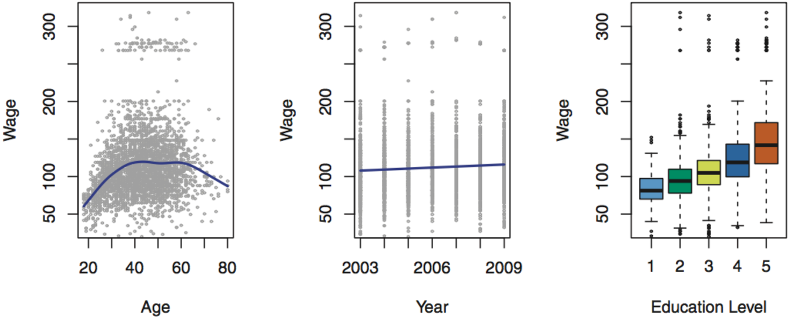
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Focus on relationships between:
  - Age and wage: Wages increase with age, then decline after age 60.
  - Year and wage: Slight wage increase between 2003 and 2009.
  - Education and wage: Higher education correlates with higher wages.



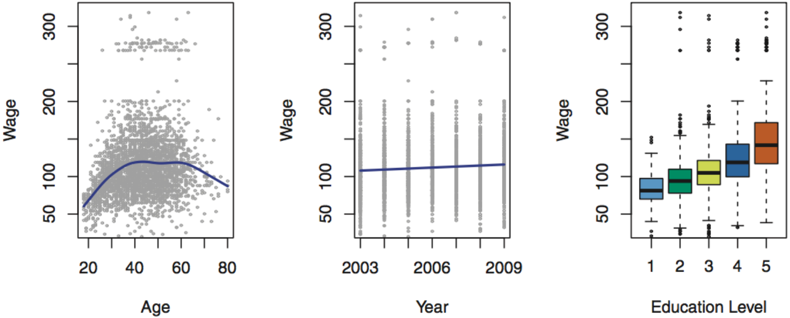
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Age alone isn't a precise predictor due to high variability.



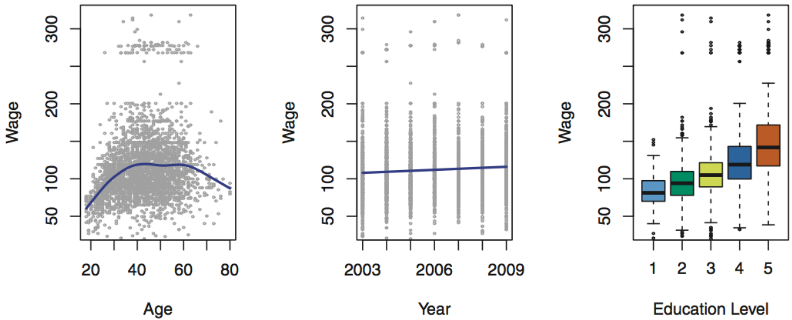
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Age alone isn't a precise predictor due to high variability.
- Best wage predictions combine age, education, and year.



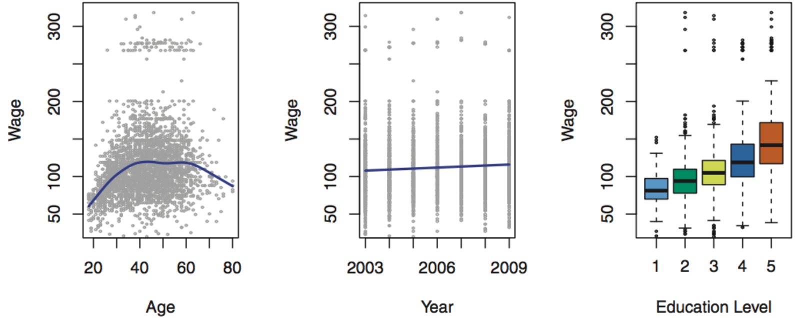
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Age alone isn't a precise predictor due to high variability.
- Best wage predictions combine age, education, and year.
- Linear regression can be used to predict wage from this data set (next chapter).



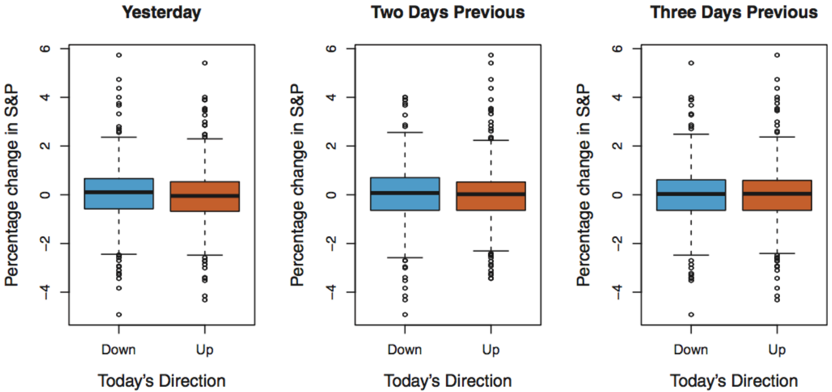
Wage Data

- **Wage data set**: Analyzes factors influencing wages for men in the Atlantic region of the U.S.
- Age alone isn't a precise predictor due to high variability.
- Best wage predictions combine age, education, and year.
- Linear regression can be used to predict wage from this data set (next chapter).
- Ideally, we should predict wage in a way that accounts for the non-linear relationship between wage and age.



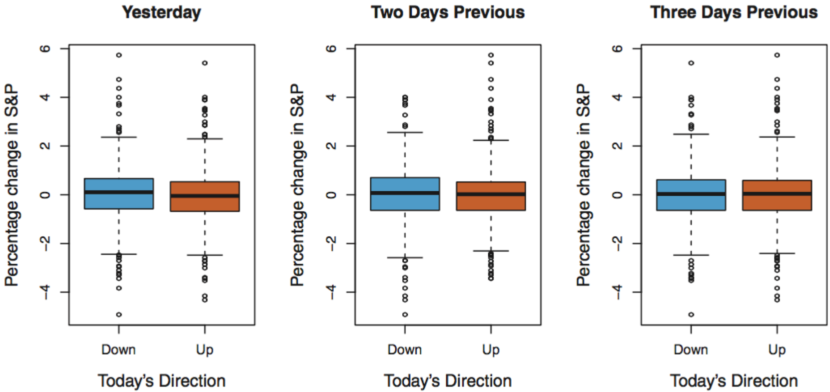
Stock Market Data

- **Smarket data set**: Analyzes daily movements of the S&P 500 index over a 5-year period.



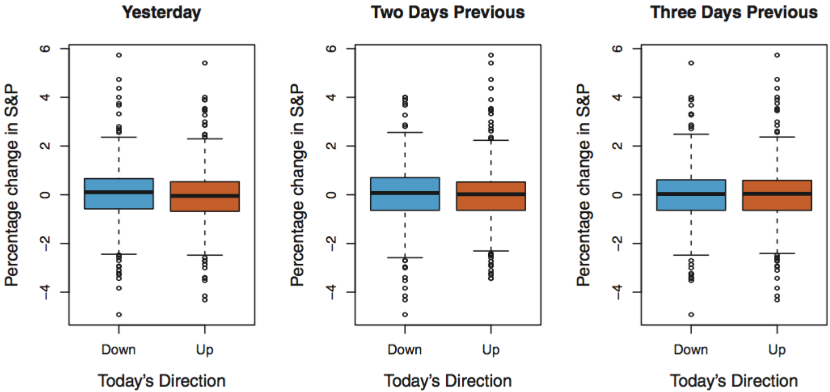
Stock Market Data

- **Smarket data set**: Analyzes daily movements of the S&P 500 index over a 5-year period.
- Goal: Predict index mvmt based on the previous days' percentage changes.



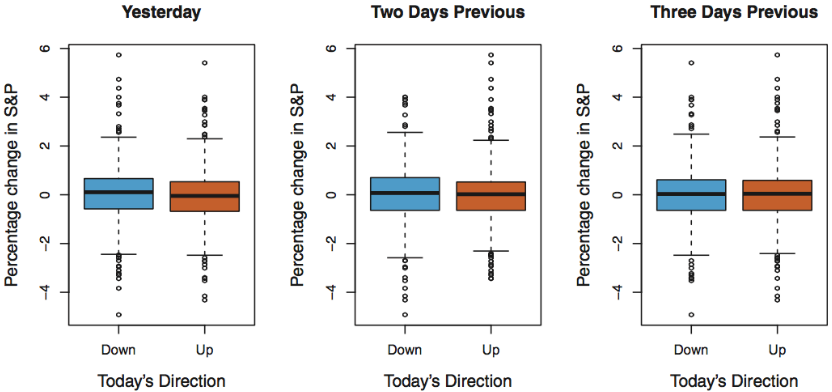
Stock Market Data

- **Smarket data set**: Analyzes daily movements of the S&P 500 index over a 5-year period.
- Goal: Predict index mvmt based on the previous days' percentage changes.
- Left-hand panel: Boxplots for the previous day's percentage change.



Stock Market Data

- **Smarket data set**: Analyzes daily movements of the S&P 500 index over a 5-year period.
- Goal: Predict index mvmt based on the previous days' percentage changes.
- Left-hand panel: Boxplots for the previous day's percentage change.
- Little difference between days the market increased vs. decreased, suggesting no simple predictive strategy.



Regression vs Classification

- **Regression Problem**:

Regression vs Classification

- **Regression Problem**:
  - Predicts a continuous or quantitative output.

Regression vs Classification

- **Regression Problem**:
  - Predicts a continuous or quantitative output.
  - Example: Wage data set where the goal is to predict wages based on factors like age and education.

Regression vs Classification

- **Regression Problem**:
  - Predicts a continuous or quantitative output.
  - Example: Wage data set where the goal is to predict wages based on factors like age and education.
- **Classification Problem**:

Regression vs Classification

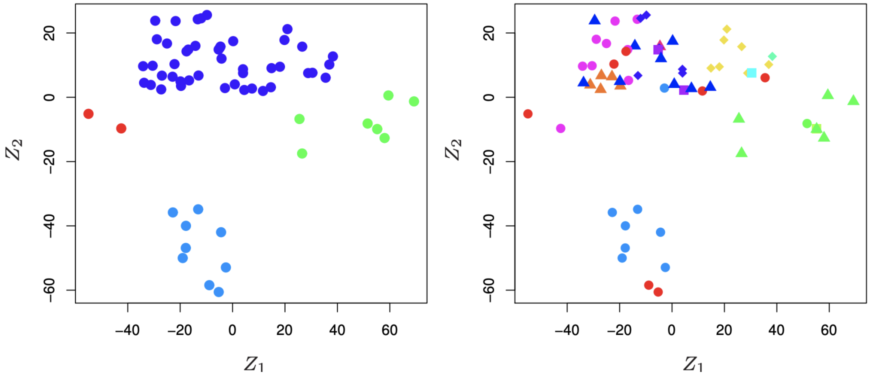
- **Regression Problem**:
  - Predicts a continuous or quantitative output.
  - Example: Wage data set where the goal is to predict wages based on factors like age and education.
- **Classification Problem**:
  - Predicts a categorical or qualitative outcome.

Regression vs Classification

- **Regression Problem**:
  - Predicts a continuous or quantitative output.
  - Example: Wage data set where the goal is to predict wages based on factors like age and education.
- **Classification Problem**:
  - Predicts a categorical or qualitative outcome.
  - Example: Smarket data set, predicting whether the S&P 500 index will increase (Up) or decrease (Down) on a given day.

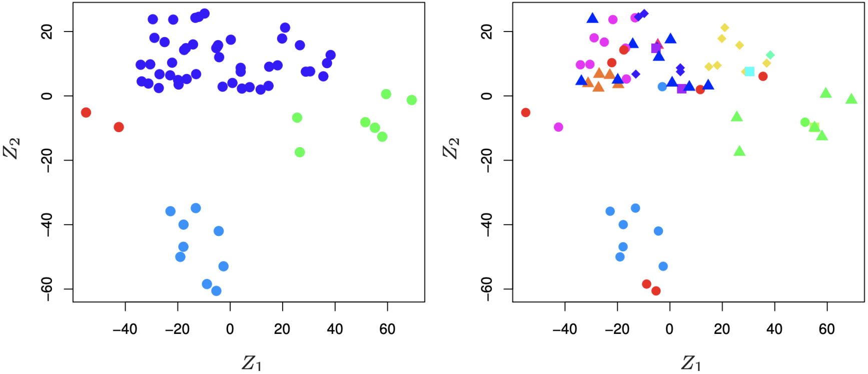
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.



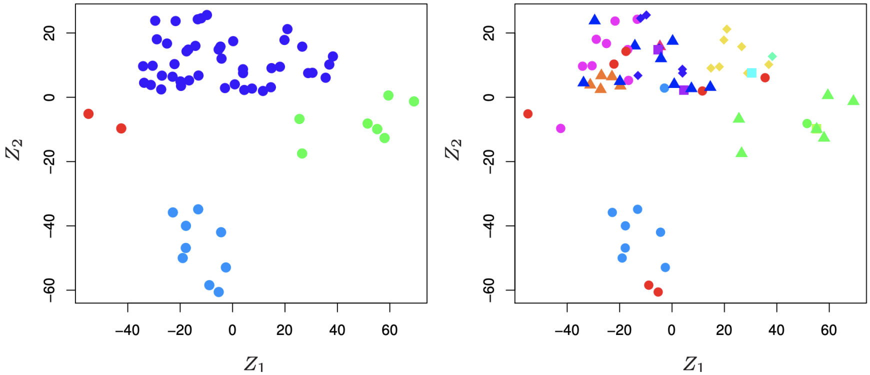
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.
- Goal: Determine clusters among cell lines based on gene expression data.



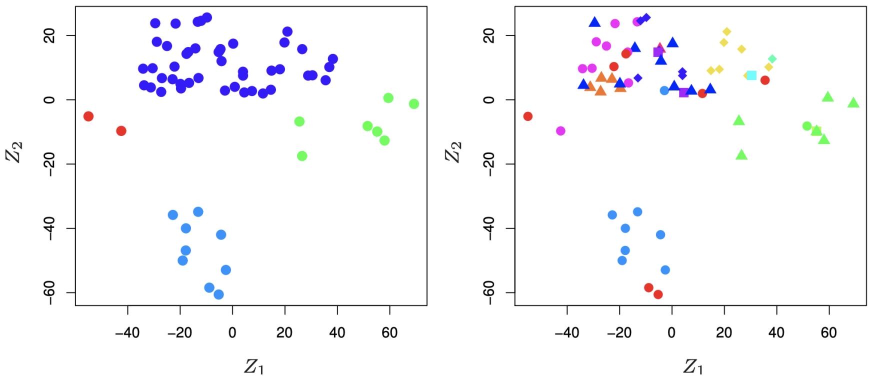
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.
- Goal: Determine clusters among cell lines based on gene expression data.
- We apply **Principal Component Analysis (PCA)**



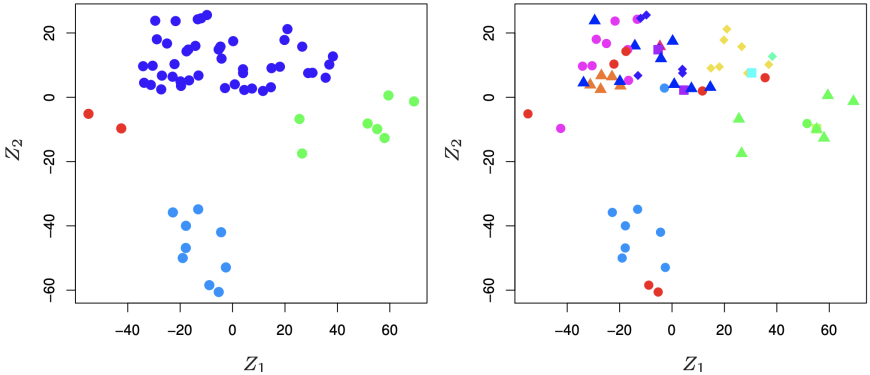
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.
- Goal: Determine clusters among cell lines based on gene expression data.
- We apply **Principal Component Analysis (PCA)**
- Reduces 6,830 measurements to two principal components (Z1 and Z2).



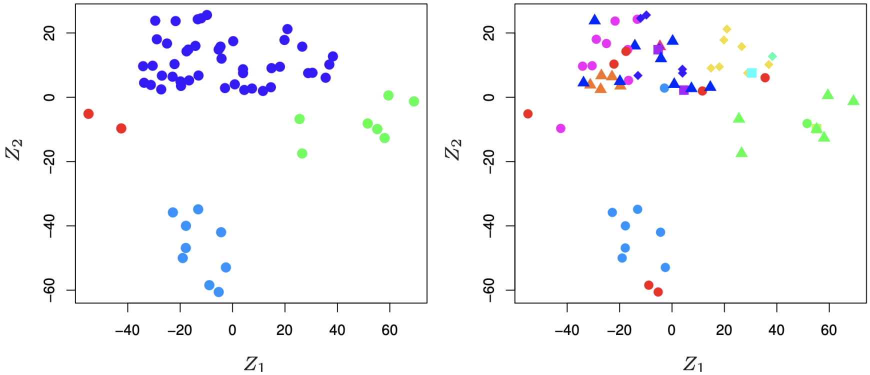
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.
- Goal: Determine clusters among cell lines based on gene expression data.
- We apply **Principal Component Analysis (PCA)**
- Reduces 6,830 measurements to two principal components (Z1 and Z2).
- Allows visualization of potential clusters in a 2D space.



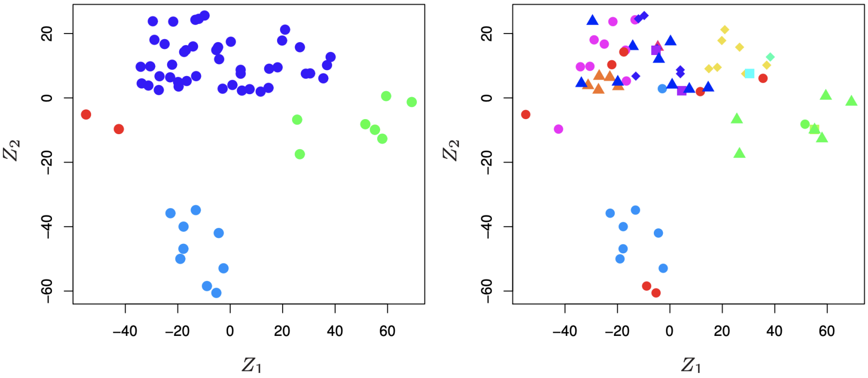
Gene Expression Data

- **NCI60 data set**: Contains 6,830 gene expression measurements for 64 cancer cell lines.
- Goal: Determine clusters among cell lines based on gene expression data.
- We apply **Principal Component Analysis (PCA)**
- Reduces 6,830 measurements to two principal components (Z1 and Z2).
- Allows visualization of potential clusters in a 2D space.
- Some information loss is expected, but clustering is now visually assessable.



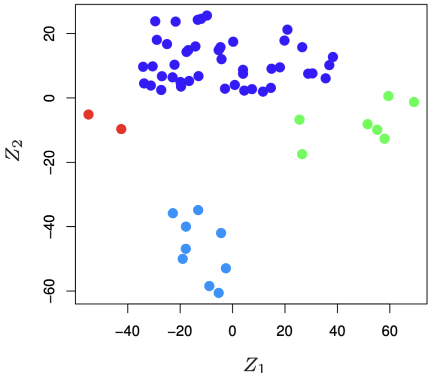
Gene Expression Data

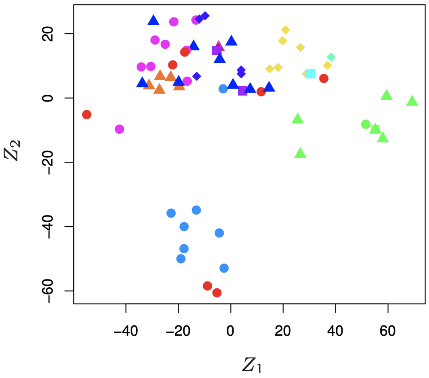
- **Left**: Rep. of the NCI60 gene expression data set in a 2D space.



Gene Expression Data

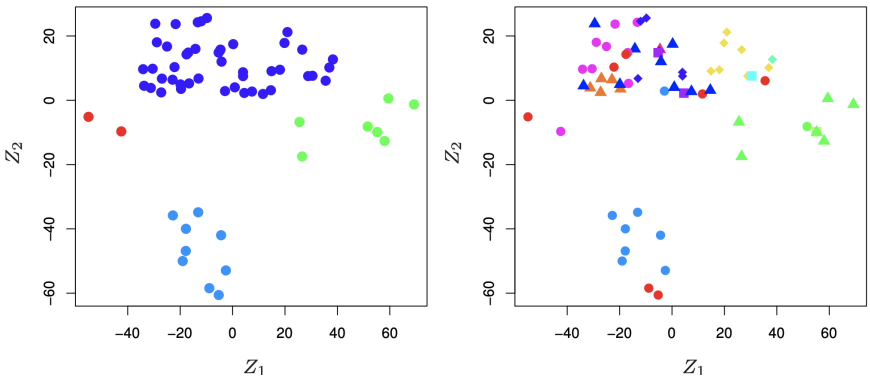
- **Left**: Rep. of the NCI60 gene expression data set in a 2D space.
- Each point corresponds to one of the 64 cell lines





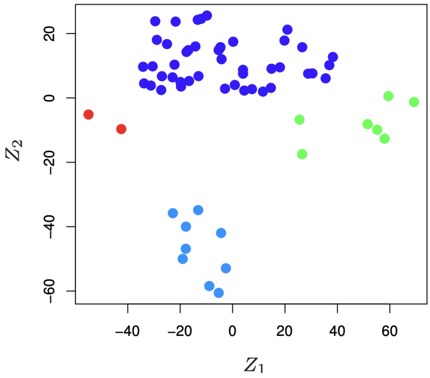
Gene Expression Data

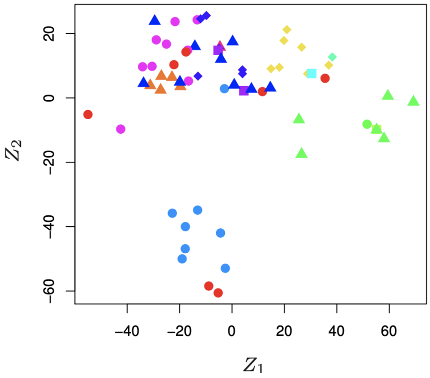
- **Left**: Rep. of the NCI60 gene expression data set in a 2D space.
- Each point corresponds to one of the 64 cell lines
- Suggests at least four clusters among the cell lines.



Gene Expression Data

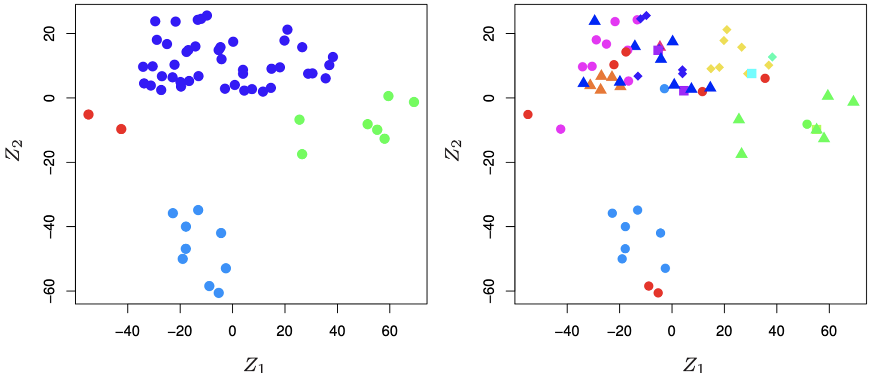
- **Left**: Rep. of the NCI60 gene expression data set in a 2D space.
- Each point corresponds to one of the 64 cell lines
- Suggests at least four clusters among the cell lines.
- **Right**: Same but displays 14 cancer types using distinct colored symbols.





Gene Expression Data

- **Left**: Rep. of the NCI60 gene expression data set in a 2D space.
- Each point corresponds to one of the 64 cell lines
- Suggests at least four clusters among the cell lines.
- **Right**: Same but displays 14 cancer types using distinct colored symbols.
- Shows that cell lines with the same cancer type are often near each other.



Supervised vs Unsupervised problems

- Wage and Smarket data sets contain both input and output variables.

Supervised vs Unsupervised problems

- Wage and Smarket data sets contain both input and output variables.
- Another important class of problems involves situations in which we only observe input variables, with no corresponding output.

Supervised vs Unsupervised problems

- Wage and Smarket data sets contain both input and output variables.
- Another important class of problems involves situations in which we only observe input variables, with no corresponding output.
- Example: Marketing setting where customers are grouped based on demographic data.

Supervised vs Unsupervised problems

- Wage and Smarket data sets contain both input and output variables.
- Another important class of problems involves situations in which we only observe input variables, with no corresponding output.
- Example: Marketing setting where customers are grouped based on demographic data.
- Goal: Identify groups or clusters of similar individuals based on observed characteristics.

Supervised learning problem

- The dataset contains n observations.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
    - In the regression problem, y_i is quantitative (e.g wage).
    - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \ldots, n$ (e.g. age, year, education level).

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \dots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \dots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \dots, (x_n, y_n)$ are usually called samples or instances.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \ldots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \ldots, (x_n, y_n)$ are usually called samples or instances.
- Based on the observed data we would like to perform:

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \ldots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \ldots, (x_n, y_n)$ are usually called samples or instances.
- Based on the observed data we would like to perform:
  - Prediction: accurately predict future outcome.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \dots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \dots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \dots, (x_n, y_n)$ are usually called samples or instances.
- Based on the observed data we would like to perform:
  - Prediction: accurately predict future outcome.
  - Estimation: understand how inputs affect the outcome.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \dots, n$.
    - In the regression problem, y_i is quantitative (e.g wage).
    - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \dots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \dots, (x_n, y_n)$ are usually called samples or instances.
- Based on the observed data we would like to perform:
    - Prediction: accurately predict future outcome.
    - Estimation: understand how inputs affect the outcome.
    - Model selection: find the best model for the outcome or which inputs affect the outcome.

Supervised learning problem

- The dataset contains n observations.
- For each observation, there is an outcome measurement, usually called dependent variable or response or target and noted y_i where $i = 1, \ldots, n$.
  - In the regression problem, y_i is quantitative (e.g wage).
  - In the classification problem, y_i takes values in a finite, unordered set (e.g. up/down).
- For each observation, there is a vector of p predictor measurements, usually called inputs or regressors or covariates or features or independent variables and denoted x_i where $i = 1, \ldots, n$ (e.g. age, year, education level).
- The observed training data $(x_1, y_1), \ldots, (x_n, y_n)$ are usually called samples or instances.
- Based on the observed data we would like to perform:
  - Prediction: accurately predict future outcome.
  - Estimation: understand how inputs affect the outcome.
  - Model selection: find the best model for the outcome or which inputs affect the outcome.
  - Inference: assess the quality of our predictions.

Unsupervised learning problem

- The dataset contains n observations.

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \ldots, n$ (e.g.gene expression measurements).

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \ldots, n$ (e.g.gene expression measurements).
- The objective is here more fuzzy. We can be interested on

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \dots, n$ (e.g.gene expression measurements).
- The objective is here more fuzzy. We can be interested on
  - Finding groups of samples that behave similarly.

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \ldots, n$ (e.g.gene expression measurements).
- The objective is here more fuzzy. We can be interested on
  - Finding groups of samples that behave similarly.
  - Find linear combinations of features with the most variation.

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \ldots, n$ (e.g.gene expression measurements).
- The objective is here more fuzzy. We can be interested on
  - Finding groups of samples that behave similarly.
  - Find linear combinations of features with the most variation.
  - Other

Unsupervised learning problem

- The dataset contains n observations.
- For each observation, there is no outcome measurement
- For each observation, there is a vector of p predictor measurements denoted x_i where $i = 1, \ldots, n$ (e.g.gene expression measurements).
- The objective is here more fuzzy. We can be interested on
  - Finding groups of samples that behave similarly.
  - Find linear combinations of features with the most variation.
  - Other
- The model assessment problem is difficult as it is hard to know how well you are doing.

Simple to sophisticated methods

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.

Simple to sophisticated methods

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.
- One has to understand the simpler methods first, in order to grasp the more sophisticated ones.

Simple to sophisticated methods

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.
- One has to understand the simpler methods first, in order to grasp the more sophisticated ones.
- It is important to accurately assess the performance of a method, to know how well or how badly it is working. Simpler methods sometimes perform as well as fancier ones.

Simple to sophisticated methods

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.
- One has to understand the simpler methods first, in order to grasp the more sophisticated ones.
- It is important to accurately assess the performance of a method, to know how well or how badly it is working. Simpler methods sometimes perform as well as fancier ones.
- This is an exciting research area, having important applications in science, industry and finance.

Simple to sophisticated methods

- It is important to understand the ideas behind the various techniques, in order to know how and when to use them.
- One has to understand the simpler methods first, in order to grasp the more sophisticated ones.
- It is important to accurately assess the performance of a method, to know how well or how badly it is working. Simpler methods sometimes perform as well as fancier ones.
- This is an exciting research area, having important applications in science, industry and finance.
- Statistical learning is a fundamental ingredient in the training of a modern data scientist.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.
- Late 1970s: more techniques for learning from data were available.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.
- Late 1970s: more techniques for learning from data were available.
- However, they were almost exclusively linear methods be- cause fitting non-linear relationships was computationally difficult at the time.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.
- Late 1970s: more techniques for learning from data were available.
- However, they were almost exclusively linear methods be- cause fitting non-linear relationships was computationally difficult at the time.
- Mid 1980s, classification and regression trees were developed, followed shortly by generalized additive models.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.
- Late 1970s: more techniques for learning from data were available.
- However, they were almost exclusively linear methods be- cause fitting non-linear relationships was computationally difficult at the time.
- Mid 1980s, classification and regression trees were developed, followed shortly by generalized additive models.
- 1980s: Neural networks gained popularity., and support vector machines arose in the 1990s.

Brief History of Statistical Learning

- Though the term statistical learning is fairly new, many of the concepts that underlie the field were developed long ago.
- Early XIXth centurey: the method of least squares was developed, implementing the earliest form of what is now known as linear regression.
- Linear regression is used for predicting quantitative values.
- 1936: In order to predict qualitative values, linear discriminant analysis was proposed.
- 1940s: various authors put forth an alternative approach, logistic regression.
- Early 1970s: the term generalized linear model was developed to describe an entire class of statistical learning methods that include both linear and logistic regression as special cases.
- Late 1970s: more techniques for learning from data were available.
- However, they were almost exclusively linear methods be- cause fitting non-linear relationships was computationally difficult at the time.
- Mid 1980s, classification and regression trees were developed, followed shortly by generalized additive models.
- 1980s: Neural networks gained popularity., and support vector machines arose in the 1990s.
- 1990s: Support vector machines arose.

Acknowledgement

We thank Professors Hastie and Tibshirani for sharing their source files for the book 'An introduction to Statistical Learning'. We also thank professor Yang Ning for sharing the material which has inspired this one.