An Introduction to Machine Learning

10S3001 – Artificial Intelligence
Samuel I. G. Situmeang
Faculty of Informatics and Electrical Engineering

Objectives

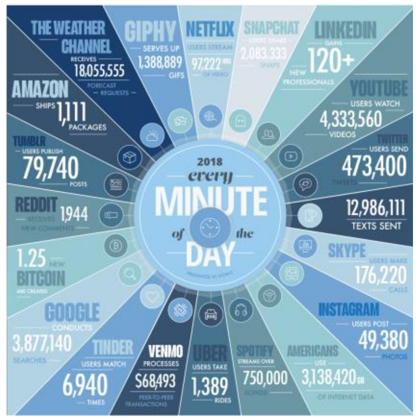
Students are able:

- to explain the fundamental concepts of machine learning, including its definition, types, and applications.
- to differentiate between supervised and unsupervised learning algorithms, providing examples of each.
- to describe the process of splitting data into training and testing sets, and explain their roles in model evaluation.
- to define overfitting and underfitting, and discuss techniques to mitigate these issues.

3

Machine Learning Concepts

Data everywhere!



Source: https://www.domo.com/

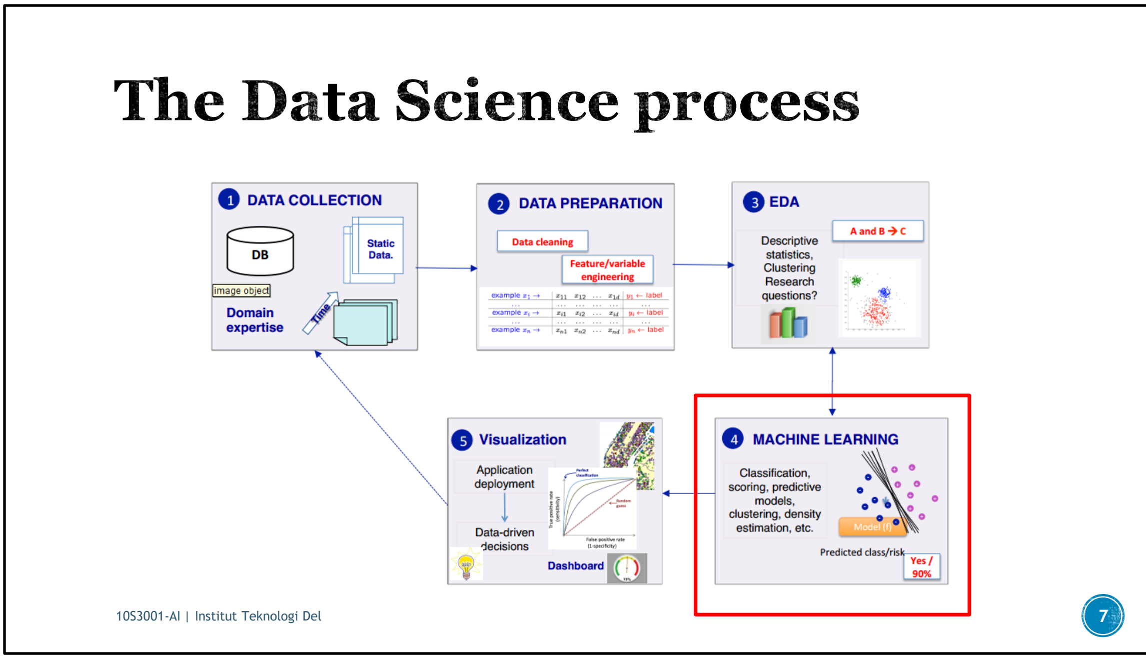
Data types

Data comes in different sizes and types:

- Texts
- Numbers
- Clickstreams
- Graphs
- Tables
- Images
- Transactions
- Videos
- Some or all of the above!

Smile, we are 'DATAFIED'!

- Wherever we go, we are "datafied".
- Smartphones are tracking our locations.
- We leave a data trail in our web browsing.
- Interaction in social networks.
- Privacy is an important issue in Data Science.



Exploratory Data Analysis (**EDA**)

Applications of ML

- We all use it on a daily basis. Examples:

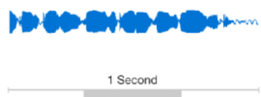
Applications of ML

- Spam filtering
- Credit card fraud detection
- Digit recognition on checks, zip codes
- Detecting faces in images
- MRI image analysis
- Recommendation system
- Search engines
- Handwriting recognition
- Scene classification
- etc...

Now is a great time to study ML

Research progress

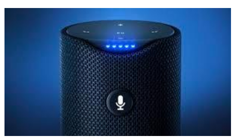
Image classification



Audio synthesis

Games

Products



Voice recognition

Google Translate

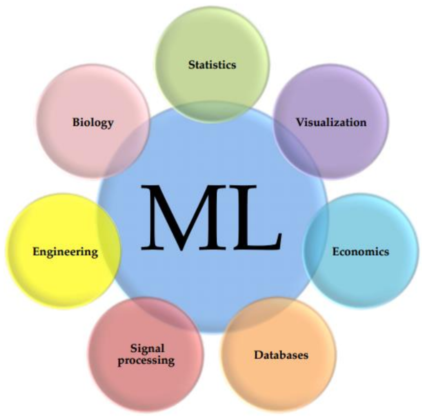
Translation

Self-driving cars

Progress in ML is driven by…

- More compute
- More data
- Better algorithms → Need more people who understand the algorithms!

Interdisciplinary field



ML versus Statistics

Statistics:

- Hypothesis testing
- Experimental design
- Analysis of variance (ANOVA)
- Linear regression
- Logistic regression
- Generalized Linear Models (GLM)
- Principal Component Analysis (PCA)

Machine Learning:

- Decision trees
- Rule induction
- Neural Networks
- Support Vector Machines (SVMs)
- Clustering method
- Association rules
- Feature selection
- Visualization
- Graphical models
- Genetic algorithm

http://statweb.stanford.edu/~jhf/ftp/dm-stat.pdf

Machine Learning definition

Alan Turing proposed the concept of a learning machine in 1950 (in the same paper that proposed the Turing test).

Idea: Divide the problem into two parts:

1. A machine that **simulates a child's brain** (analogous to a blank notebook: should function by simple mechanisms and have lots of blank sheets).
2. A way of **teaching the child machine** (should be simple since we know how to teach a human child).

Teacher rewards good behaviour and penalizes bad behaviour.

Machine Learning definition

"An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside."

Alan Turing

- While we don't know *how* our brain converts input to output, we know what the output should be for every input.
- We can use this knowledge to teach the machine.

Machine Learning definition

"How do we create computer programs that improve with experience?"

Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

Machine Learning definition

"How do we create computer programs that improve with experience?"

Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Tom Mitchell. Machine Learning 1997.

Machine Learning definition

- A branch of **artificial intelligence**, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.
- As intelligence requires knowledge, it is necessary for computers to acquire knowledge.

Cabang kecerdasan buatan, berkaitan dengan desain dan pengembangan algoritma yang memungkinkan komputer untuk mengembangkan perilaku berdasarkan data empiris.

Types of machine learning Algorithms

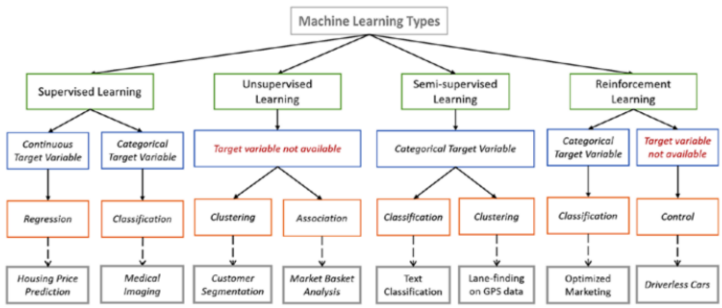
There some variations of how to define the types of Machine Learning Algorithms but commonly they can be divided into categories according to their purpose and the main categories are the following:

- **Supervised learning** (predictive model, "labeled" data).
  - Classification
  - Numeric prediction/forecasting/regression
- **Unsupervised learning** (descriptive model, "unlabeled" data).
  - Clustering
  - Pattern Discovery
- **Semi-supervised learning** (mixture of "labeled" and "unlabeled" data).
- **Reinforcement learning**. Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions.

Common algorithms for each categories

- Supervised learning
  - Classification. e.g. Logistic Regression, Decision Tree, KNN, Random Forest, SVM, & Naive Bayes
  - Numeric prediction/forecasting/regression. e.g. Linear Regression, KNN, Gradient Boosting & AdaBoost
- Unsupervised learning
  - Clustering. e.g. K-Means
  - Pattern Discovery. e.g. Apriori, FP-Growth, & Eclat
- Semi-supervised learning
- Reinforcement learning.
  - e.g. Q-Learning, Temporal Difference (TD), & Deep Adversarial Networks

A brief of ML types with sample use cases



Source: https://en.proft.me/

22 Supervised Learning and Unsupervised Learning

Supervised Learning

Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n) / x_i \in \mathbb{R}^d$ and y_i is the label.

example $x_1 \rightarrow$	x_{11} x_{12} … x_{1d}	$y_1 \leftarrow$ label
…	… … … …	…
example $x_i \rightarrow$	x_{i1} x_{i2} … x_{id}	$y_i \leftarrow$ label
…	… … … …	…
example $x_n \rightarrow$	x_{n1} x_{n2} … x_{nd}	$y_n \leftarrow$ label

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
…				
fruit n	…	…	…	…

Supervised Learning

Training data: "examples" x with "labels" y.

$$(x_1, y_1), \dots, (x_n, y_n) / x_i \in \mathbb{R}^d$$

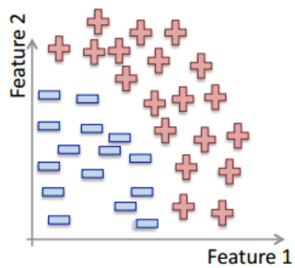
- **Classification**: y is discrete. To simplify, $y \in \{-1. +1\}$

$f: \mathbb{R}^d \to \{-1, +1\}$ $\quad$ f is called a binary classifier

Example: Approve credit yes/no, spam/ham, banana/orange.

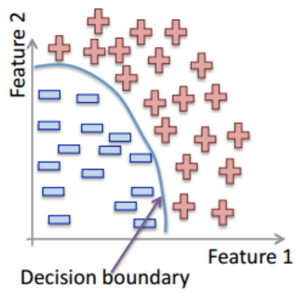
Supervised Learning

Classification:



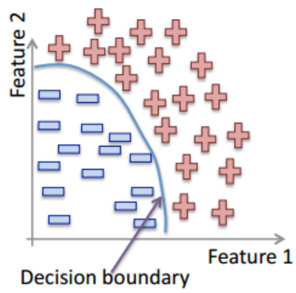
Supervised Learning

Classification:



Supervised Learning

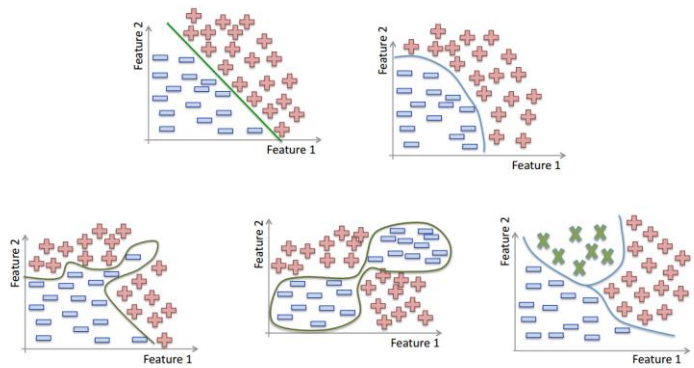
Classification:



Methods: Support Vector Machines, neural networks, decision trees, K-nearest neighbors, naive Bayes, etc.

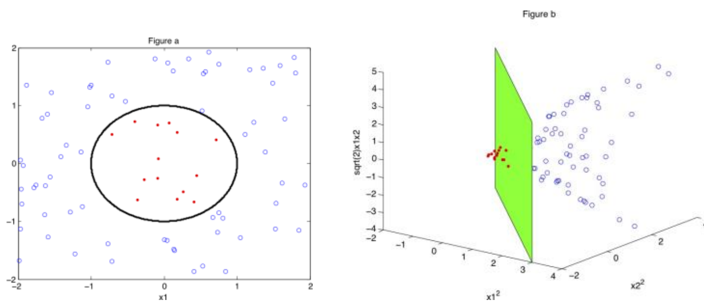
Supervised Learning

Classification:



Supervised Learning

Non linear classification



Supervised Learning

Training data: "examples" x with "labels" y.

$$(x_1, y_1), \dots, (x_n, y_n) / x_i \in \mathbb{R}^d$$

- **Regression**: y is a real value, $y \in \mathbb{R}$

$$f: \mathbb{R}^d \to \mathbb{R} \quad f \text{ is called a regressor}$$

Example: amount of credit, weight of fruit.

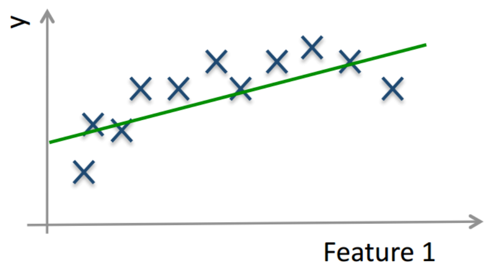
Supervised Learning

Regression:

Example: Income in function of age, weight of the fruit in function of its length.

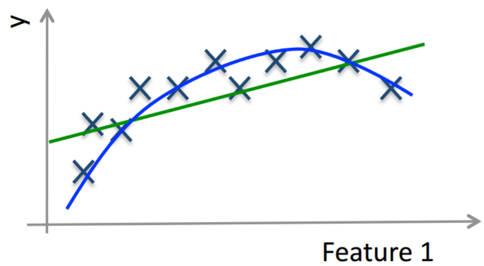
Supervised Learning

Regression:



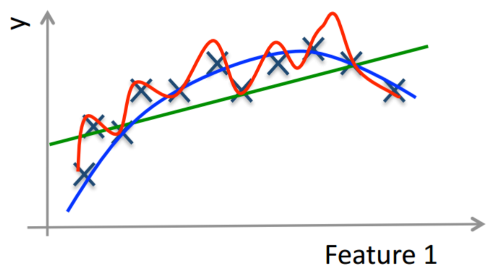
Supervised Learning

Regression:



Supervised Learning

Regression:



Unsupervised Learning

Training data: "examples" x.

$$x_1, \ldots, x_n, x_i \in X \subset \mathbb{R}^n$$

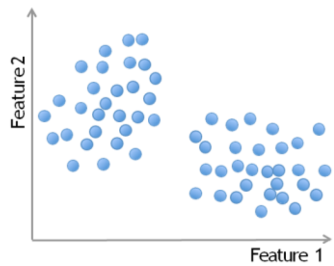
- **Clustering/segmentation**:

$$f: \mathbb{R}^d \rightarrow \{C_1, \ldots, C_k\} \quad \text{(set of cluster)}$$

Example: Find clusters in the population, fruits, species.

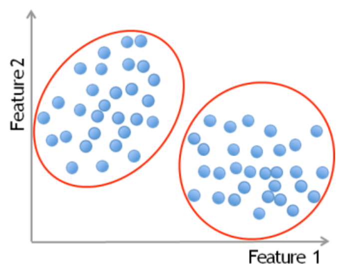
Unsupervised Learning

Clustering/segmentation:



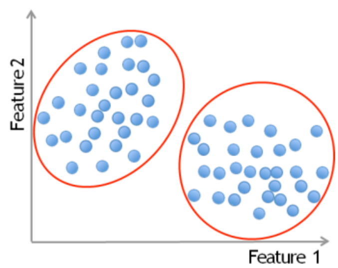
Unsupervised Learning

Clustering/segmentation:



Unsupervised Learning

Clustering/segmentation:



Methods: K-means, gaussian mixtures, hierarchical clustering, spectral clustering etc.

39 Training-Testing Concepts

Training and Testing

Training set

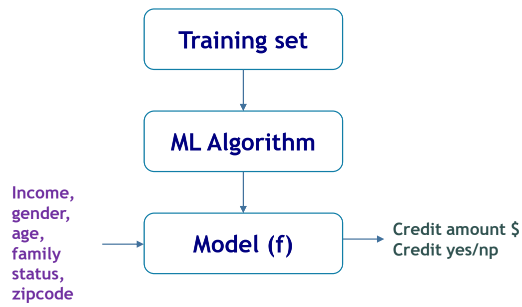
↓

ML Algorithm

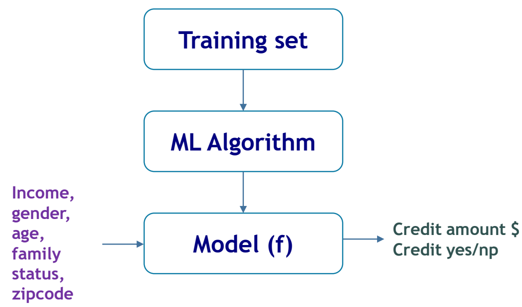
↓

Model (f)

Training and Testing



Training and Testing



Question: How can we be confident about f?

Training and Testing

- We calculate E^{train} the in-sample error (training error or empirical error/risk).

$$E^{train}(f) = \sum_{i=1}^{n} loss(y_i, f(x_i))$$

prediction label

true label

Tells us how many errors is our model, f doing on the training data itself

Training and Testing

- We calculate E^{train} the in-sample error (training error or empirical error/risk).

$$E^{train}(f) = \sum_{i=1}^{n} loss\big(y_i, f(x_i)\big)$$

- Examples of loss functions:
  - Classification error:

$$loss\big(y_i, f(x_i)\big) = \begin{cases} 1 & sign(y_i) \neq sign(f(x_i)) \\ 0 & \text{otherwise} \end{cases}$$

Training and Testing

- We calculate E^{train} the in-sample error (training error or empirical error/risk).

$$E^{train}(f) = \sum_{i=1}^{n} loss\big(y_i, f(x_i)\big)$$

- Examples of loss functions:
  - Classification error:

$$loss\big(y_i, f(x_i)\big) = \begin{cases} 1 & sign(y_i) \neq sign(f(x_i)) \\ 0 & \text{otherwise} \end{cases}$$

  - Least square loss:

$$loss\big(y_i, f(x_i)\big) = \big(y_i - f(x_i)\big)^2$$

Training and Testing

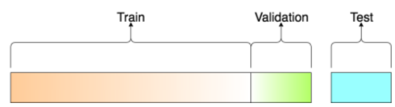
- We calculate E^{train} the in-sample error (training error or empirical error/risk).

$$E^{train}(f) = \sum_{i=1}^{n} loss(y_i, f(x_i))$$

- We aim to have $E^{train}(f)$ small, i.e., minimize $E^{train}(f)$
- We hope that $E^{train}(f)$, the out-sample error (test/true error), will be small too.

Train, Validation and Test

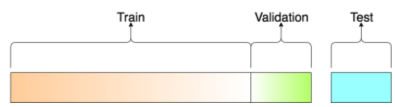
Example: Split the data randomly into 60% for training, 20% for validation and 20% for testing.



Source: https://towardsdatascience.com/

Train, Validation and Test

Training set is a set of examples used for learning a model (e.g., a classification model).

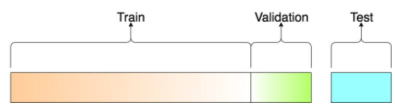


Source: https://towardsdatascience.com/

Train, Validation and Test

Training set is a set of examples used for learning a model (e.g., a classification model).

Validation set is a set of examples that cannot be used for learning the model but can help tune model parameters (e.g., selecting K in K-NN). Validation helps control overfitting.



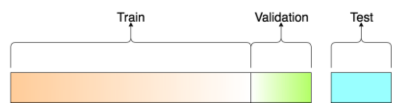
Source: https://towardsdatascience.com/

Train, Validation and Test

Training set is a set of examples used for learning a model (e.g., a classification model).

Validation set is a set of examples that cannot be used for learning the model but can help tune model parameters (e.g., selecting K in K-NN). Validation helps control overfitting.

Test set is used to assess the performance of the final model and provide an estimation of the test error.



Source: https://towardsdatascience.com/

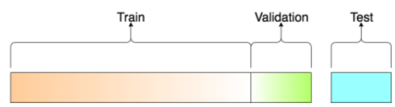
Train, Validation and Test

Training set is a set of examples used for learning a model (e.g., a classification model).

Validation set is a set of examples that cannot be used for learning the model but can help tune model parameters (e.g., selecting K in K-NN). Validation helps control overfitting.

Test set is used to assess the performance of the final model and provide an estimation of the test error.

Note: Never use the test set in any way to further tune the parameters or revise the model.



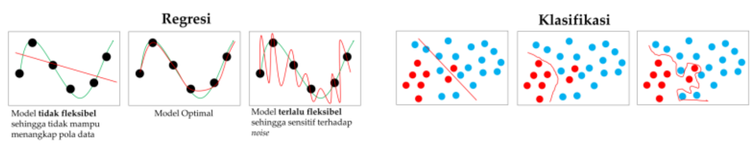
Source: https://towardsdatascience.com/

52

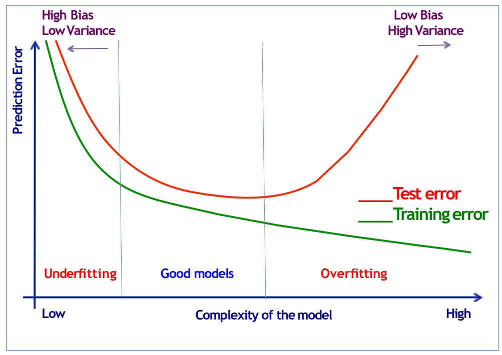
Overfitting and Underfitting

Overfitting and Underfitting

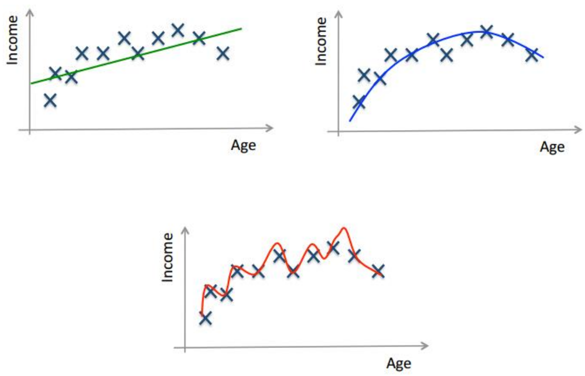
- **Overfitting**: keadaan ketika model memiliki kinerja baik hanya untuk *training data/seen examples* tetapi tidak memiliki kinerja baik untuk *unseen examples*.
  - Terjadi ketika model terlalu fleksibel (memiliki kemampuan yang terlalu tinggi untuk mengestimasi banyak fungsi) atau terlalu mencocokkan diri terhadap training data.
- **Underfitting**: keadaan ketika model memiliki kinerja buruk baik untuk *training data* dan *unseen examples*.
  - Terjadi akibat model yang telalu tidak fleksibel (memiliki kemampuan yang rendah untuk mengestimasi variasi fungsi).



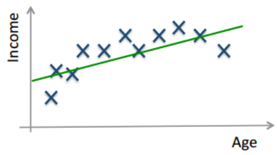
Structural Risk Minimization



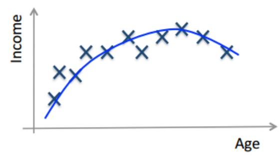
Training and Testing



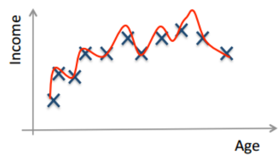
Training and Testing



High bias (underfitting)



Just right!



High variance (overfitting)

Avoid overfitting

- In general, use simple models!
  - **Reduce the number** of features manually or do feature selection.
  - Do a **model selection**.
  - Use **regularization** (keep the features but reduce their importance by setting small parameter values).
  - Do a **cross-validation** to estimate the test error.

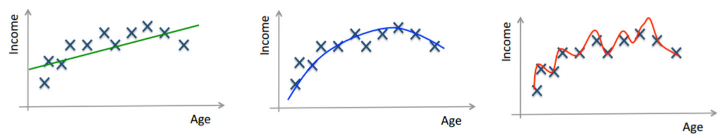
Regularization: Intuition

We want to minimize:

$$\text{Classification term} + C \times \text{Regularization term}$$

$$\sum_{i=1}^{n} loss\big(y_i, f(x_i)\big) + C \times R(f)$$

Regularization: Intuition



$f(x) = \lambda_0 + \lambda_1 x \; \ldots \; (1)$

$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 \; \ldots \; (2)$

$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4 \; \ldots \; (3)$

Hint: Avoid high-degree polynomials.

K-fold Cross Validation

A method for estimating test error using training data.

Algorithm:

Given a learning algorithm $\mathcal{A}$ and a dataset $\mathcal{D}$

Step 1: Randomly partition D into k equal-size subsets $D1; : : : ; Dk$
Step 2:

For $j = 1$ *to* k

Train $\mathcal{A}$ on all D_i, $i \in 1, \ldots, k$ and $i \neq j$, and get f_j

Apply f_j to D_j and compute E^{D_j}

Step 3: Average error over all folds.

$$\sum_{j=1}^{k} (E^{D_j})$$

Terminology review

Review the concepts and terminology:

Instance, example, feature, label, supervised learning, unsupervised learning, classification, regression, clustering, prediction, training set, validation set, test set, K-fold cross validation, classification error, loss function, overfitting, underfitting, regularization.

Machine Learning Books

1. Tom Mitchell, Machine Learning.
2. Abu-Mostafa, Yaser S. and Magdon-Ismail, Malik and Lin, Hsuan-Tien, Learning From Data, AMLBook.
3. The elements of statistical learning. Data mining, Inference, and Prediction T. Hastie, R. Tibshirani, J. Friedman.
4. Christopher Bishop. Pattern Recognition and Machine Learning.
5. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern Classification. Wiley

Machine Learning Resources

- Major journals/conferences: ICML, NIPS, UAI, ECML/PKDD, JMLR, MLJ, etc.
- Machine learning video lectures:
  http://videolectures.net/Top/Computer_Science/Machine_Learning/
- Machine Learning (Theory):
  http://hunch.net/
- LinkedIn ML groups: \Big Data" Scientist, etc.
- Women in Machine Learning:
  https://groups.google.com/forum/#!forum/women-in-machine-learning
- KDD nuggets
  http://www.kdnuggets.com/

References

- S. J. Russell and P. Borvig, *Artificial Intelligence: A Modern Approach (4th Edition)*, Prentice Hall International, 2020.
  - Chapter 19. Learning from Examples
- T. Mitchell, *Machine Learning*, 1997.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*, 2009.

eof