COSI 104a Introduction to machine learning

Chapter 4 – Classification

Instructor: Dr. Hongfu Liu
Email: hongfuliu@brandeis.edu

Main source: https://www-users.cs.umn.edu/~kumar001/dmbook/index.php

Outline

- **Pipeline**
- **Evaluation**

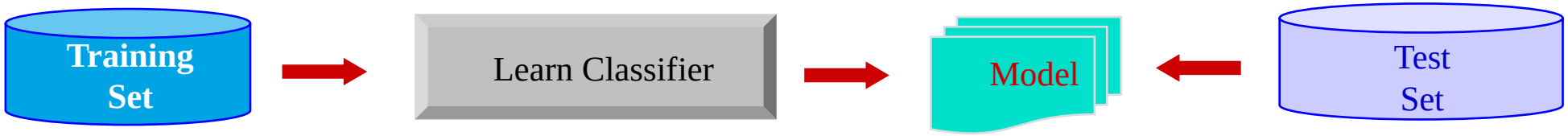
Classification Task

- **Find a model for class attribute as a function of the values of other attributes for predicting new data**

class

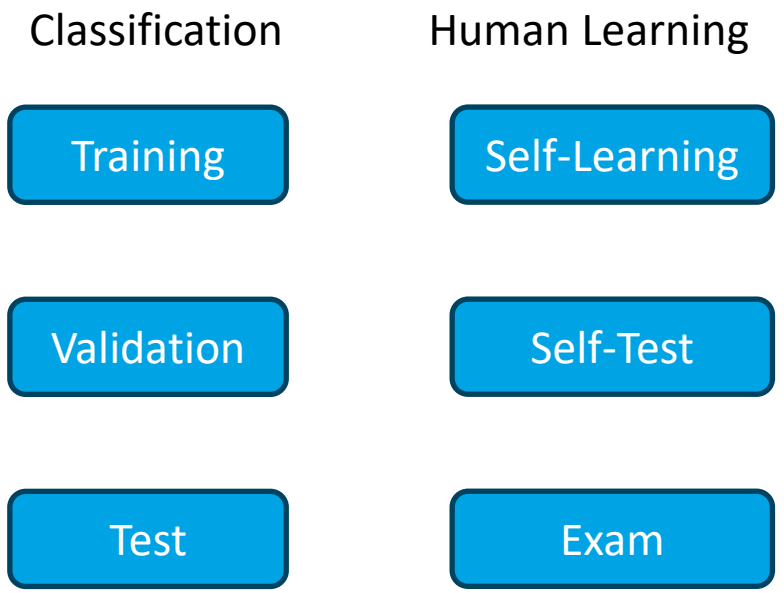
Tid	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Graduate	5	Yes
2	Yes	High School	2	No
3	No	Undergrad	1	No
4	Yes	High School	10	Yes
…	…	…	…	…

Tid	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Undergrad	7	?
2	No	Graduate	3	?
3	Yes	High School	2	?
…	…	…	…	…



Three Sets in Classification

- Training
  - Complete Sample (feature + label)
- Validation
  - Complete Sample (feature + label)
- Test
  - Incomplete Sample (only feature)



The goal of classification is to achieve **high performance** on test data

How and when ML works

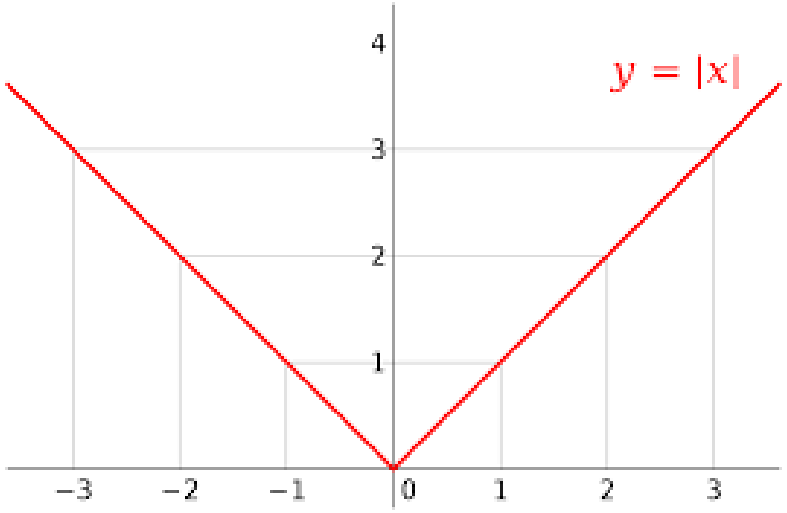
- Assumption vs Hypothesis
- What is the truth?
- The truth is just luck.
- The harder you work, the luckier you get.

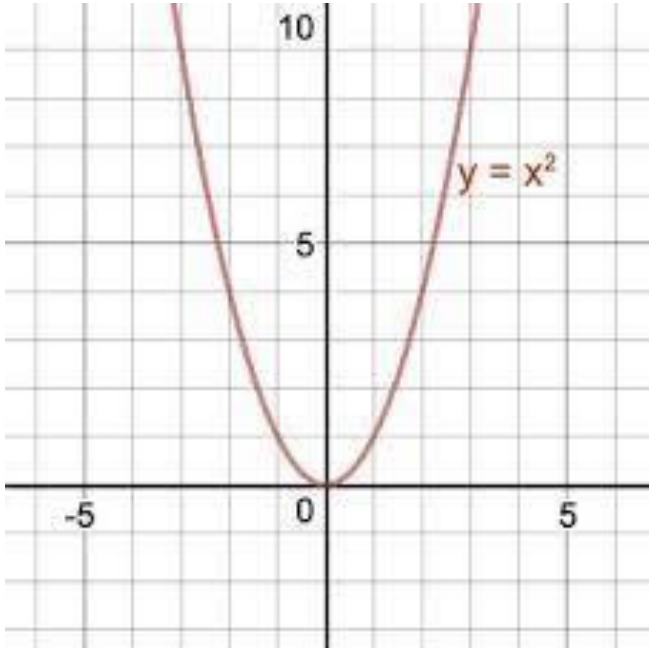
Model Parameter

- What is an optimization problem?

Optimization

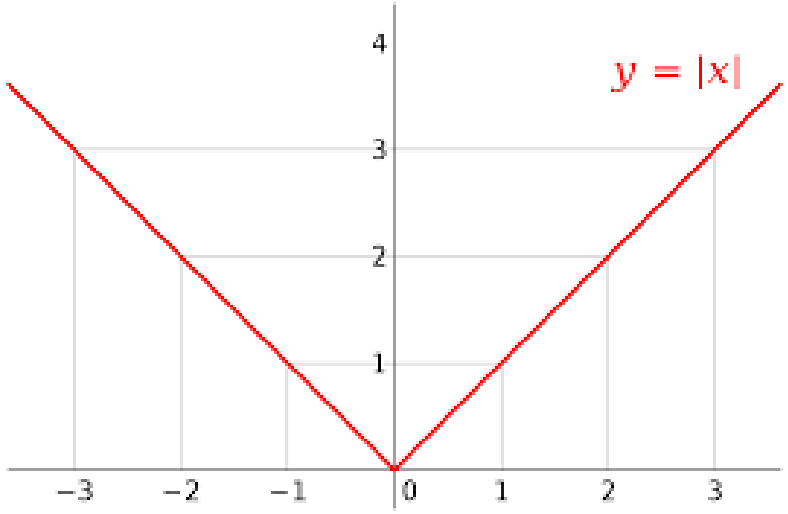
- Key elements in optimization
  - Criterion (Objective function + direction)
  - Variables
  - Constraint (Solution space)

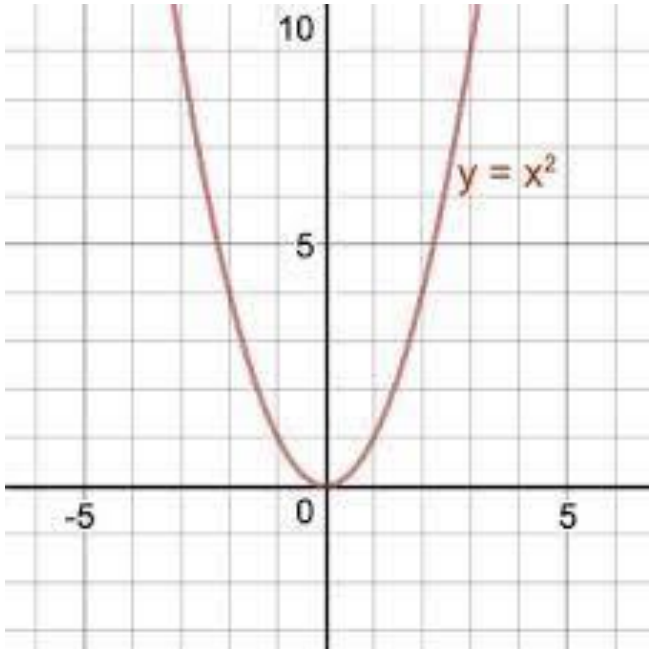




Linear Programming

- Linear programming is a special case of optimization, where the largest degree of all the variables is 1 and the solution space is convex.

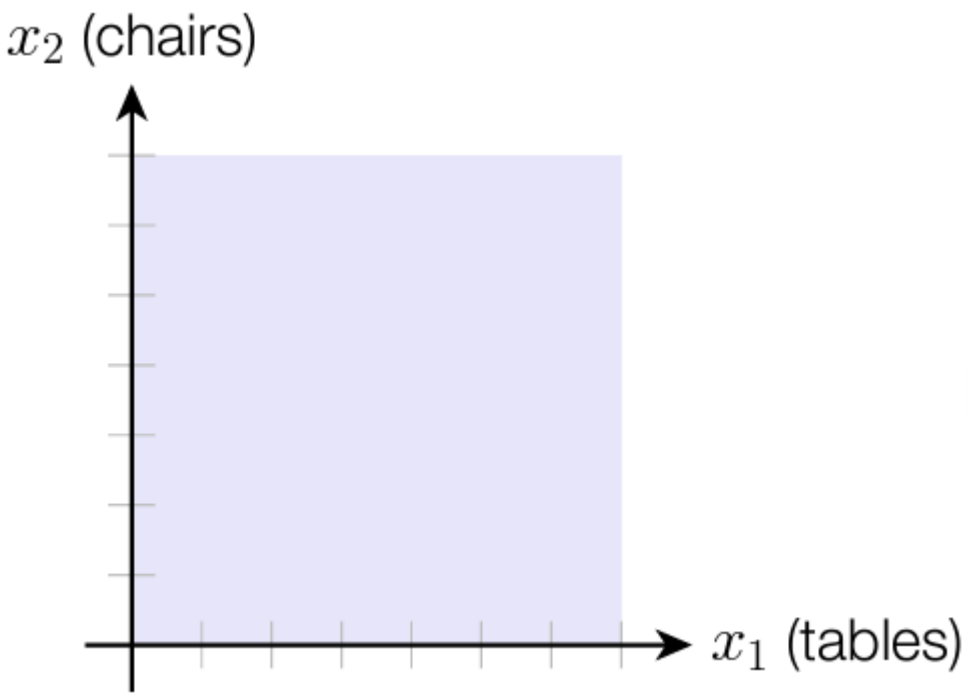


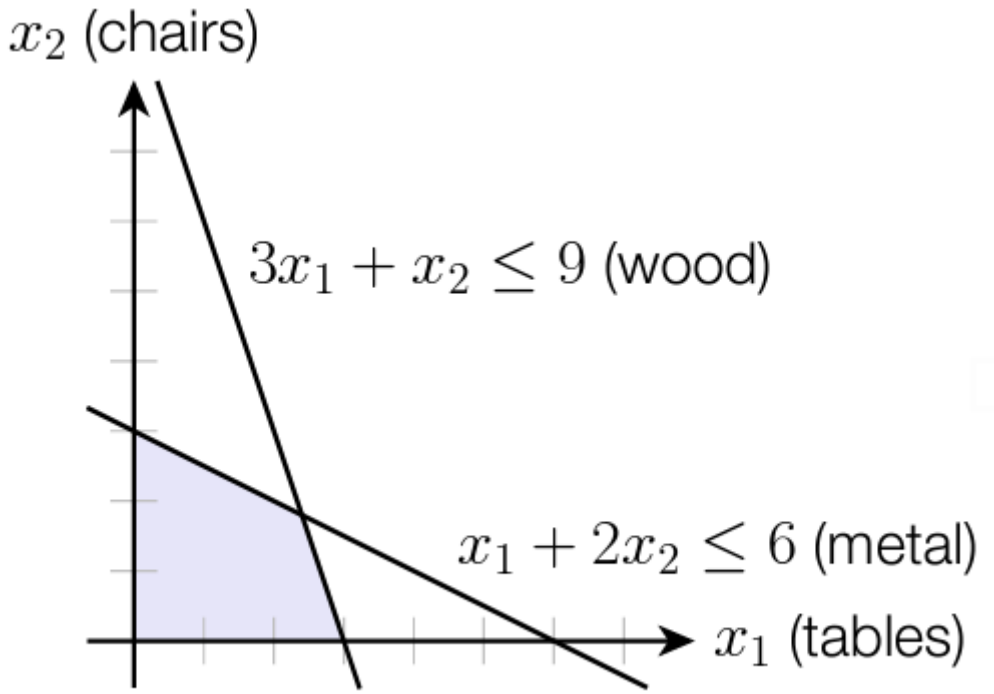


Example

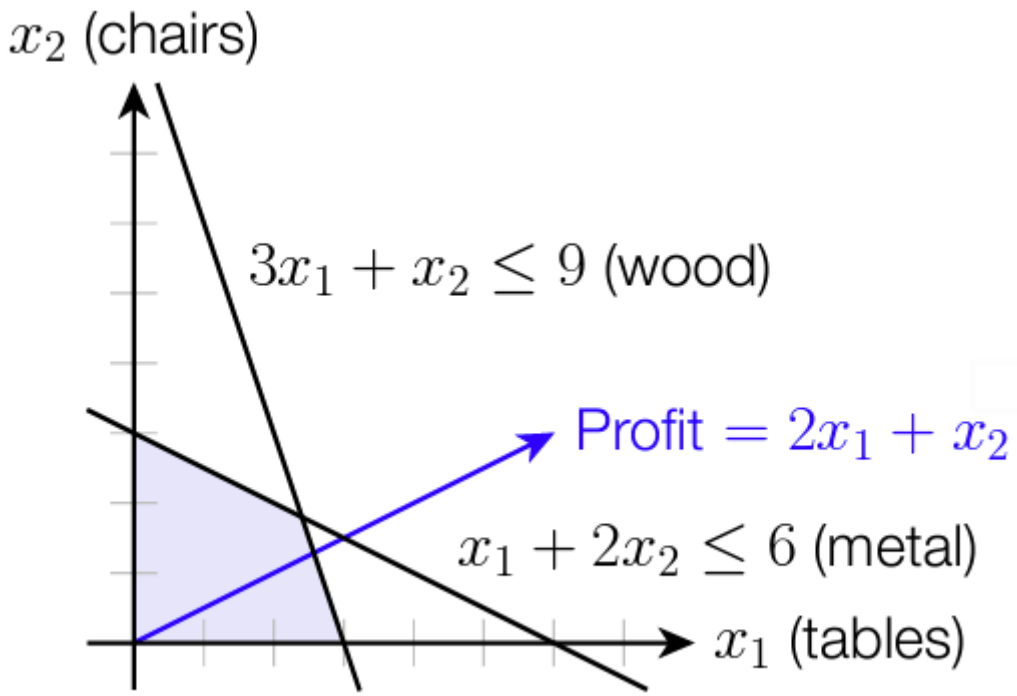
- A large factory makes tables and chairs. Each table returns a profit of $200 and each chair a profit of $100. Each table takes 1 unit of metal and 3 units of wood and each chair takes 2 units of metal and 1 unit of wood. The factory has 6K units of metal and 9K units of wood. How many tables and chairs should the factory make to maximize profit?

Example


x_2 (chairs)
x_1 (tables)


x_2 (chairs)
$3x_1 + x_2 \leq 9$ (wood)
$x_1 + 2x_2 \leq 6$ (metal)
x_1 (tables)

Example



convex

Example

- A large factory makes tables and chairs. Each table returns a profit of $200 and each chair a profit of $100. Each table takes 1 unit of metal and 3 units of wood and each chair takes 2 units of metal and 1 unit of wood. The factory has 6K units of metal and 9K units of wood. How many tables and chairs should the factory make to maximize profit?

$$\begin{aligned} \underset{x_1, x_2}{\text{maximize}} \quad & 2x_1 + x_2 \\ \text{subject to} \quad & x_1 + 2x_2 \leq 6 \\ & 3x_1 + x_2 \leq 9 \\ & x_1, x_2 \geq 0 \end{aligned}$$

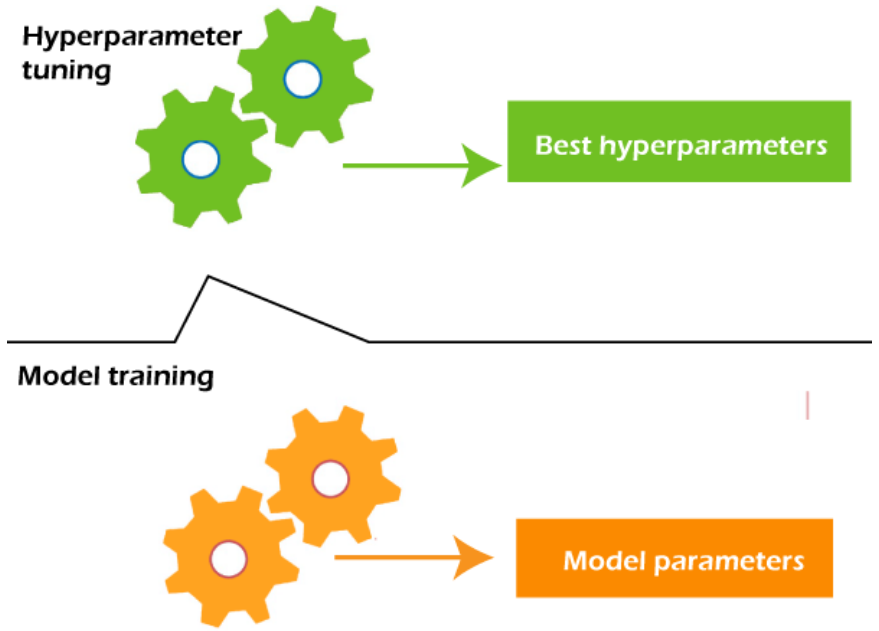
- Key elements in optimization
  - Criterion (Objective function + direction)
  - Variables
  - Constraint (Solution space)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

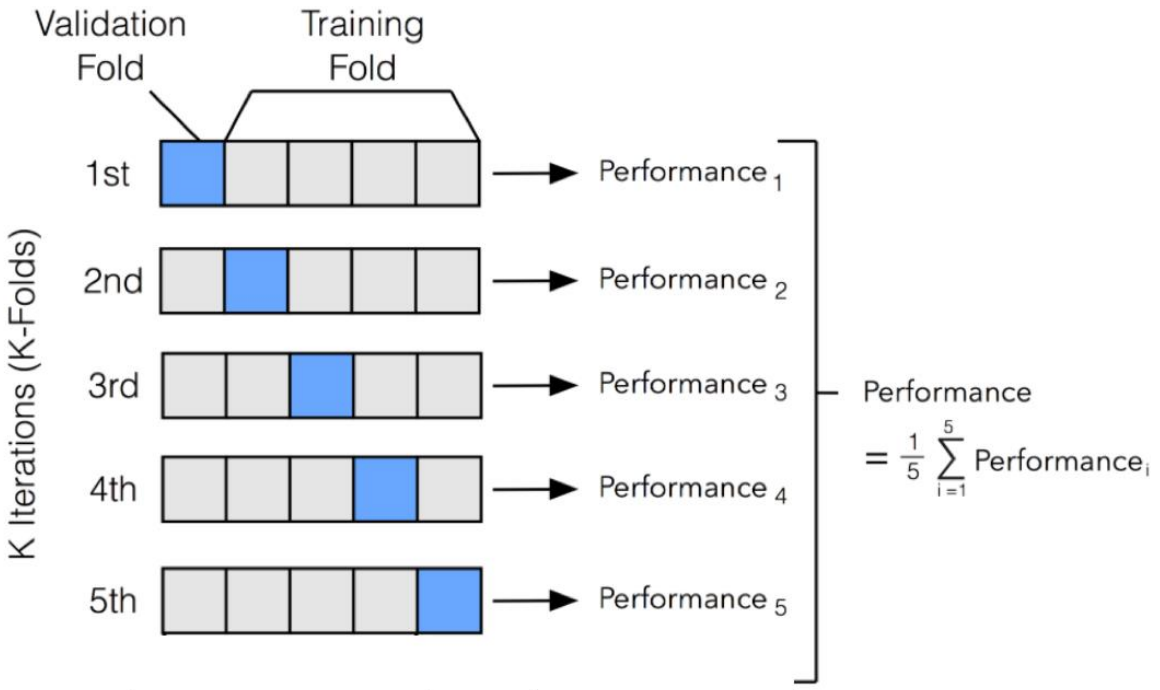
Model Parameter VS Hyperparameter

- What is model parameter?

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\frac{1}{N}\sum_i \ell(x_i,y_i)+\frac{\lambda}{2}\|\theta\|_2^2.$$



Cross-validation



Evaluation

Confusion Matrix		Ground Truth	
		P – Positive	*N* – Negative
Prediction	Positive	*TP* – True Positive	*FP* – False Positive
	Negative	*FN* – False Negative	*TN* – True Negative

Accuracy = $\frac{TP+TN}{P+N}$

Sensitivity (or recall, true positive rate) = $\frac{TP}{P}$

Specificity (or selectivity, true negative rate) = $\frac{TN}{N}$

Precision (or positive predictive value) = $\frac{TP}{TP+FP}$

False positive rate (or fall-out) = $\frac{FP}{N}$ = 1 – specificity

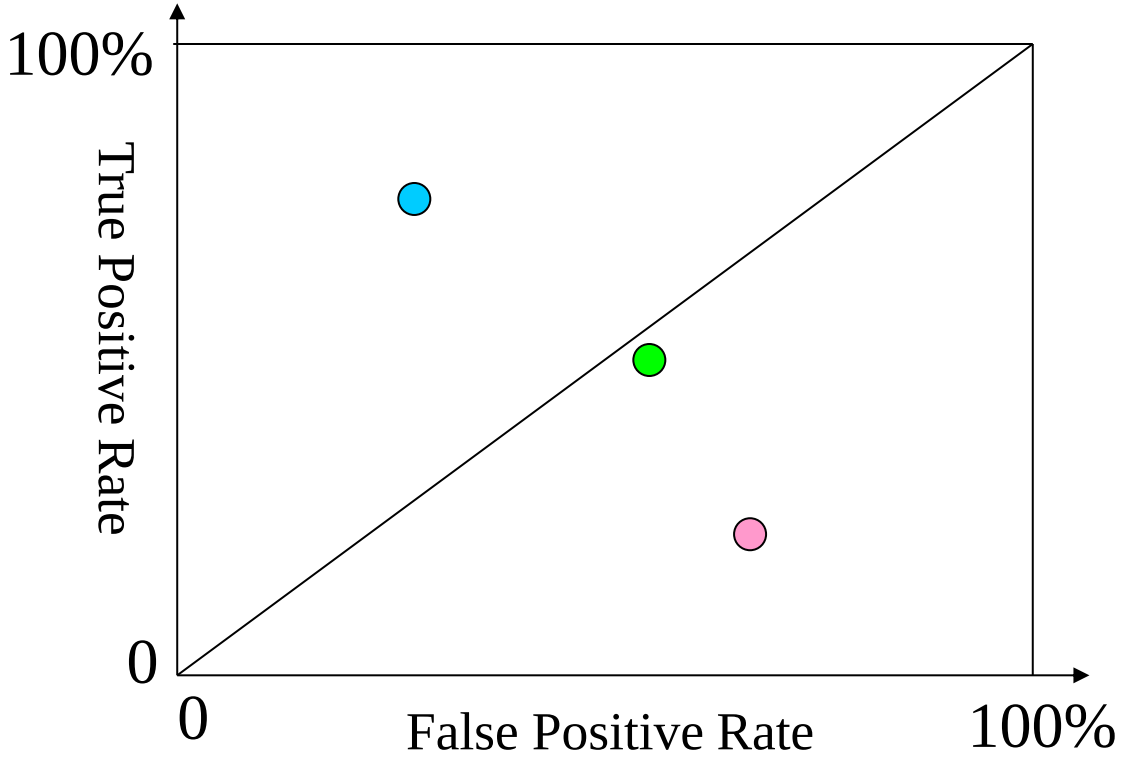
False negative rate (or miss rate) = $\frac{FN}{P}$ = 1 – sensitivity

False discovery rate (FDR) = $\frac{FP}{TP+FP}$ = 1 – Precision

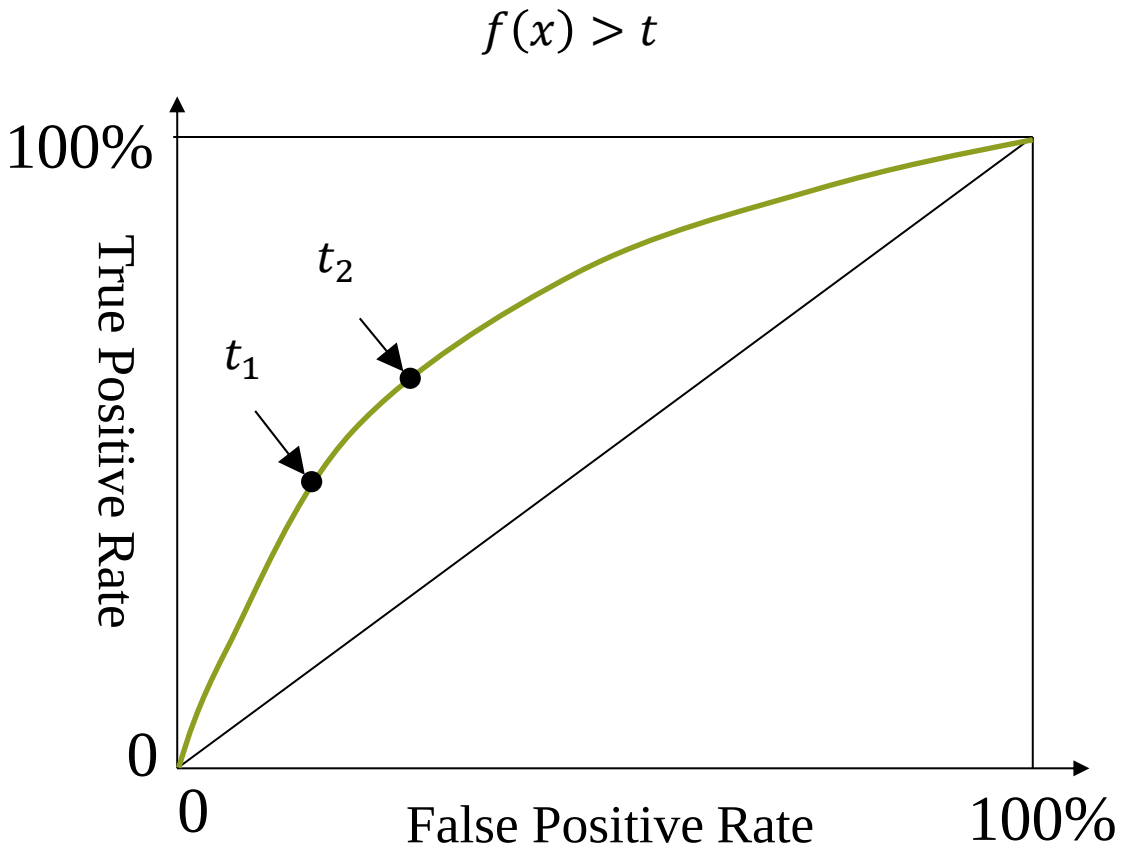
$$\mathrm{F}_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

$$\mathrm{F}_\beta = \frac{1+\beta^2}{\frac{1}{precision} + \frac{\beta^2}{recall}}$$

ROC – Receiver Operating Characteristic

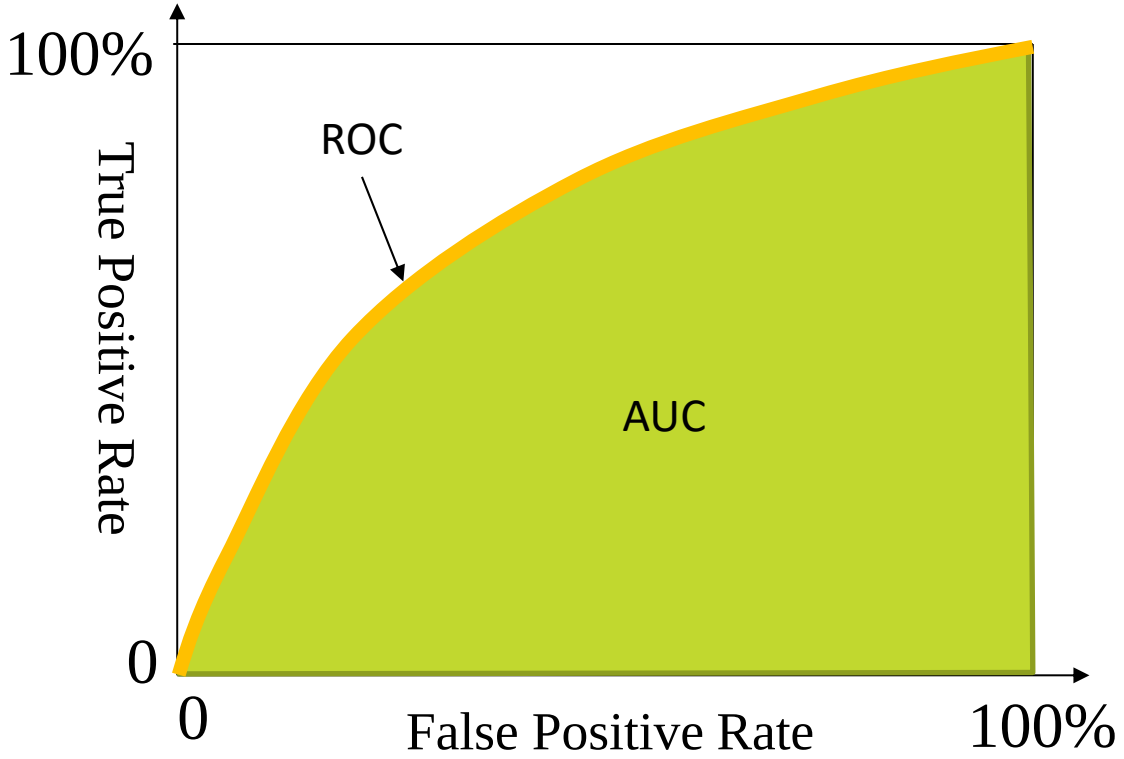


ROC Curve



AUC – Area Under Curve

Wikipedia: “The Area Under Curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. … …. The machine learning community most often uses the ROC AUC statistic for model comparison. However, this practice has recently been questioned based upon new machine learning research that shows that the AUC is quite noisy as a classification measure and has some other significant problems in model comparison.”


100%
True Positive Rate
ROC
AUC
0
0
False Positive Rate
100%

Summary

- Goal of classification
- Three sets
- Model parameter vs hyperparameter
- Cross-validation
- Evaluation


?
Q
&
&
A
!