What is Machine Learning?

Deep Learning Course – 2021
E. Fatemizadeh

Machine Learning - Definition

- Machine learning (ML) is the scientific study of ***algorithms*** and ***statistical models*** that computer systems use to perform a specific task ***without*** using ***explicit instructions***, relying on ***patterns*** and ***inference*** instead. (From Wiki, from ML Journal) → *learn from data*
- How?
  - Build a ***mathematical model*** based on sample data, known as "***training data***",
- Related Area (Subset or superset):
  - Artificial Intelligence, Data Mining, Pattern Recognition, Computational Intelligence, Computational Statistics, Statistical Learning, Expert System, ...

Two Major Tasks

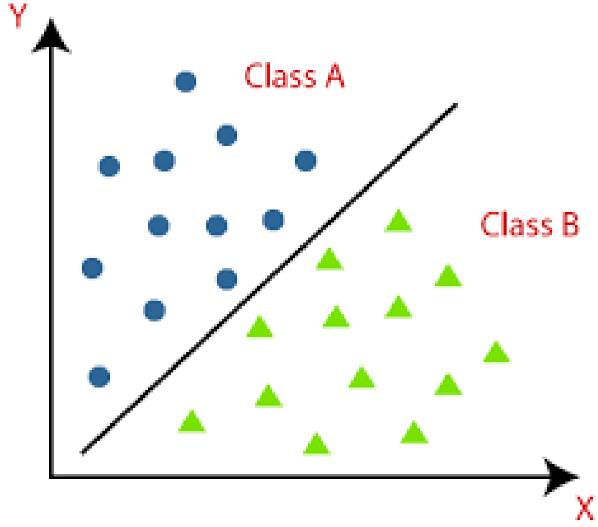
- Supervised Learning:
  - Learn Model from both **input**(s) and **output**(s) data
  - **Input**(s): Object Properties
  - **Output**(s): Discrete or “Limited Set” (e.g. Labels) or Continuous
- Unsupervised Learning:
  - Not desired output!
  - Find structure of data/Clustering/Grouping/…
- Related field:
  - Active learning, Reinforcement learning, Semi-Supervised Learning, Vector Quantization, …

Supervised Learning

› Classification:

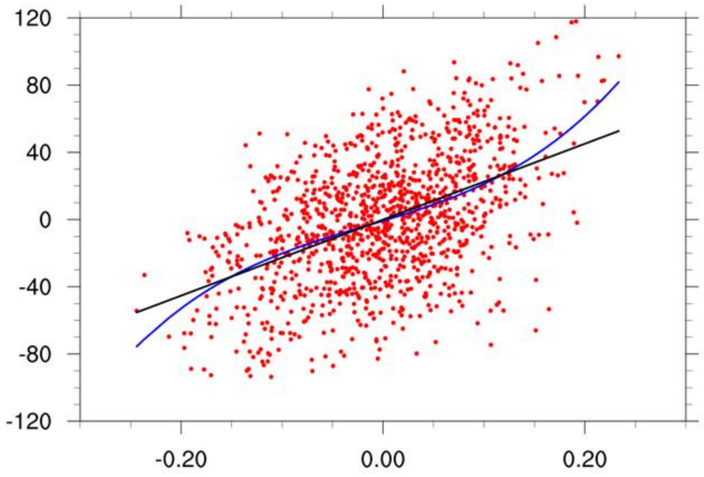
  – Input(s) are objects: Electric Signals/Image/Speech/DNA sequences/Email/….

  – Output(s) are binary (mostly Labels): “benign”/”malignant” (Tumor), “Primary”, “Social”, “Promotion” (**G**mail),



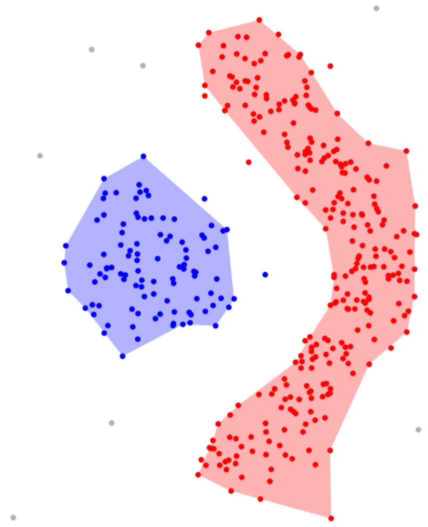
Supervised Learning:

- Regression:
  - Input(s) are objects: Electric Signals/Image/Speech/DNA sequences/Email/….
  - Output(s) are Continuous: Temperature, Score, Fuel Consumption, …



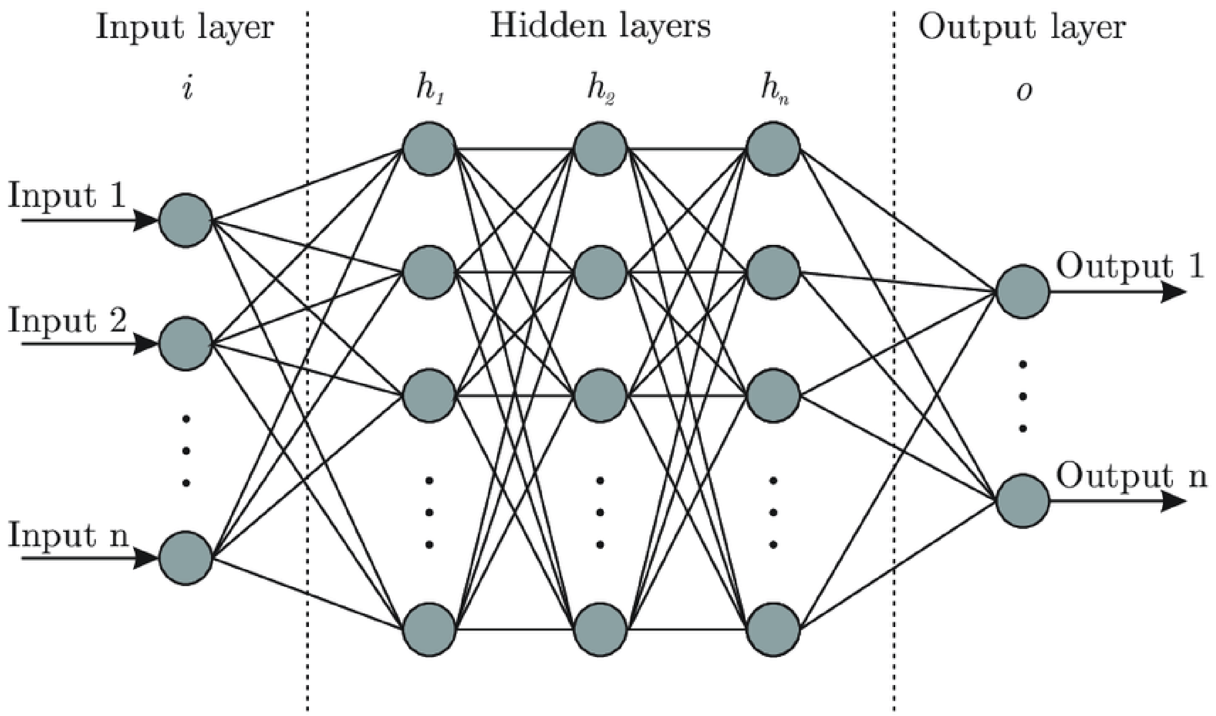
Unsupervised Learning

- Clustering/Structure Finding:
  - Input(s) are objects:
    - Electric Signals/Image/Speech/DNA sequences/Email/....
  - No Output data!



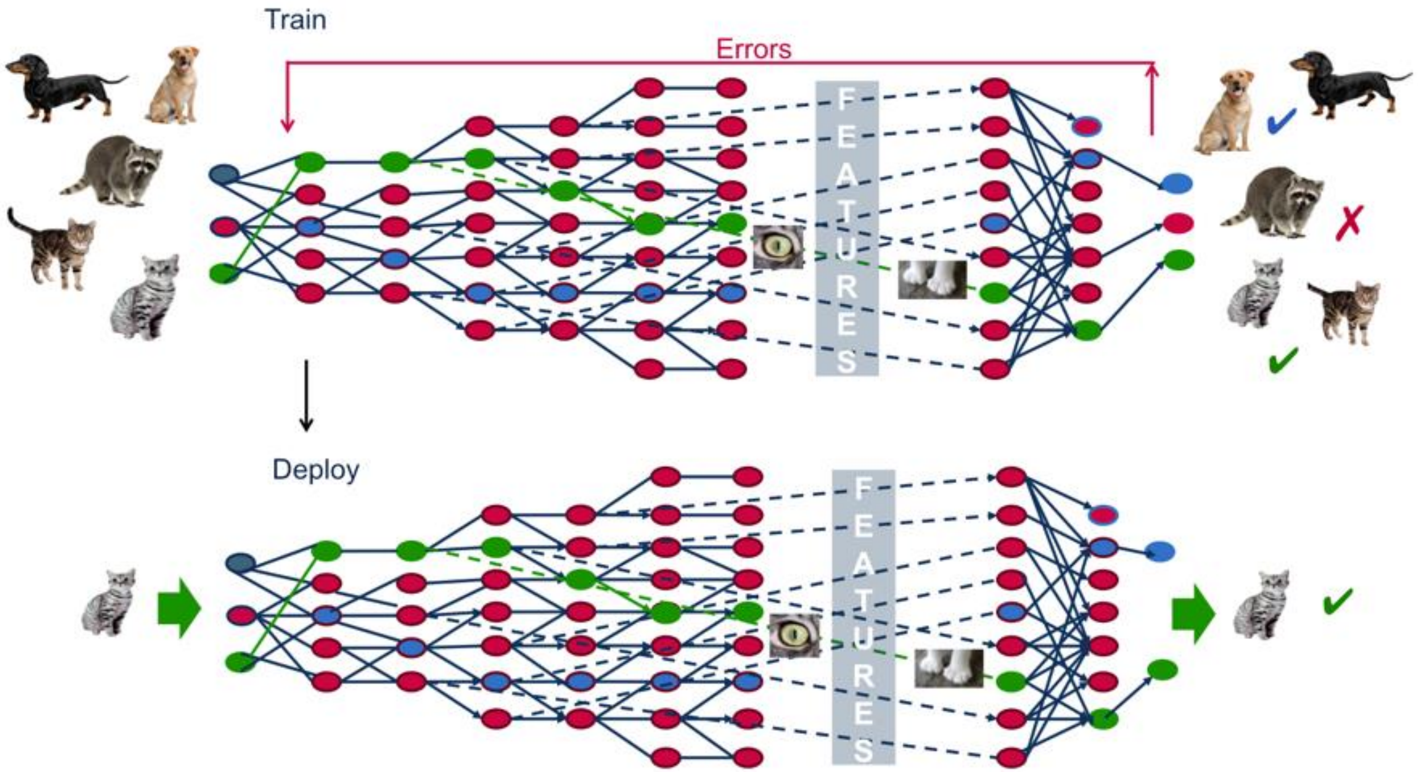
Supervised Learning Models

- Artificial Neural Networks (Shallow):
  - Biologically inspired Algorithm



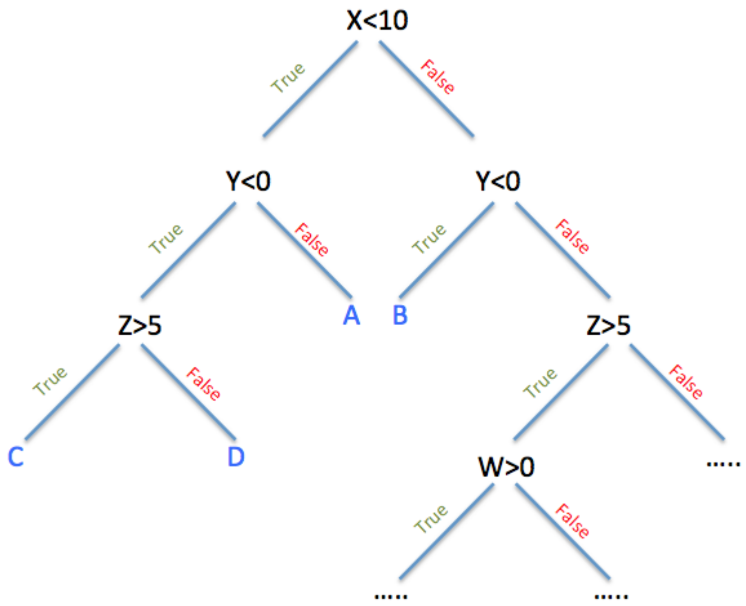
Supervised Learning Models

› Deep Artificial Neural Networks:



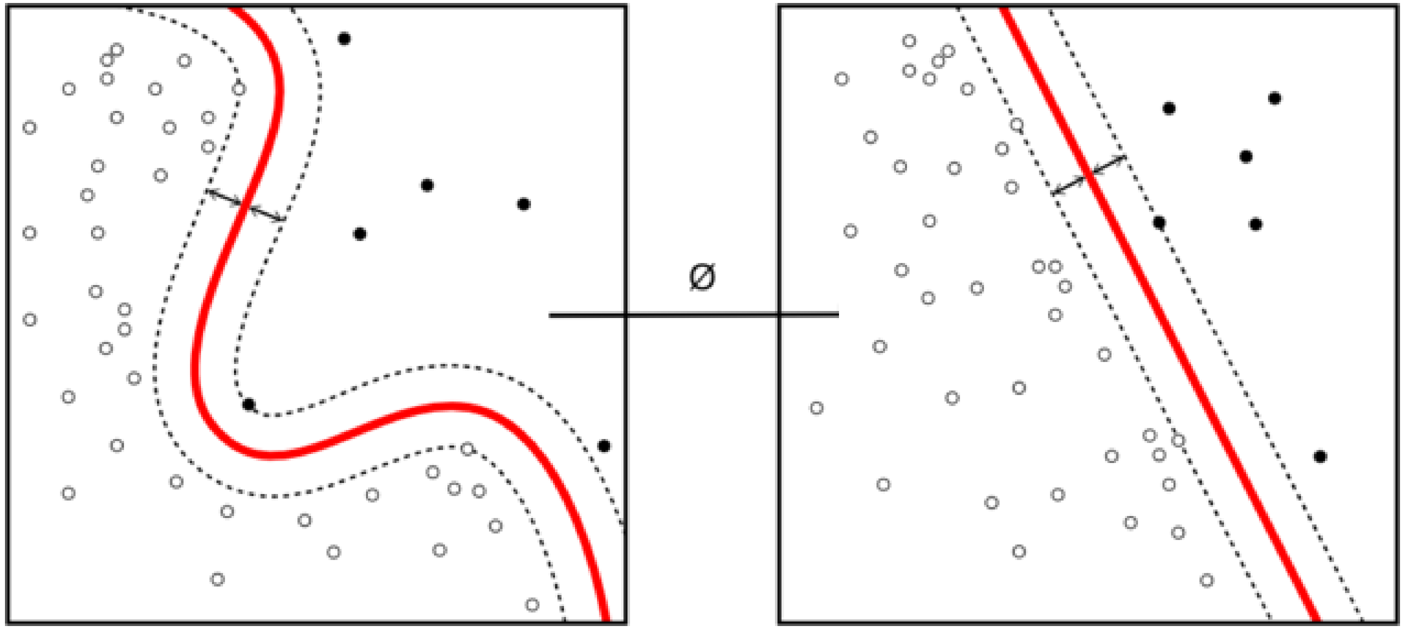
Supervised Learning Models

› Decision Trees:
  – A Huge set of if-then rules in tree-line structure, which build systematically!



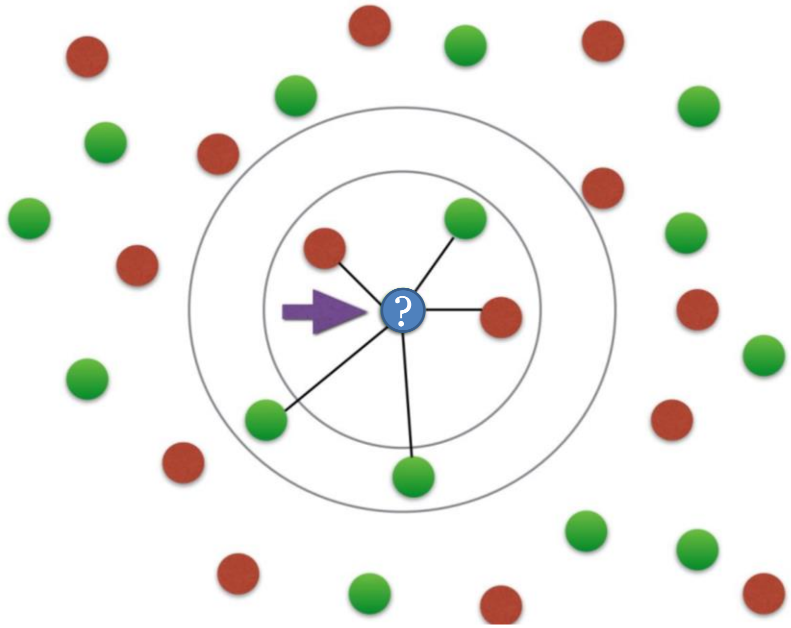
Supervised Learning Models

› Support Vector Machines (SVM):
  – A margin based classifier/regressor:



Supervised Learning Models

- K-NN (k-nearest neighbors algorithm):
  - An object is classified by a plurality vote of its neighbors



3-NN (Small Circle)
5-NN (Large Circle)

Supervised Learning Models

- Bayes Classifier (Optimal):
- For M-Classes with available posterior probabilities:

$$\text{Assign } \boldsymbol{x} \text{ to } \omega_i = \arg\max_{\omega_j} P(\omega_j|\boldsymbol{x}), \quad j = 1, 2, \ldots, M.$$

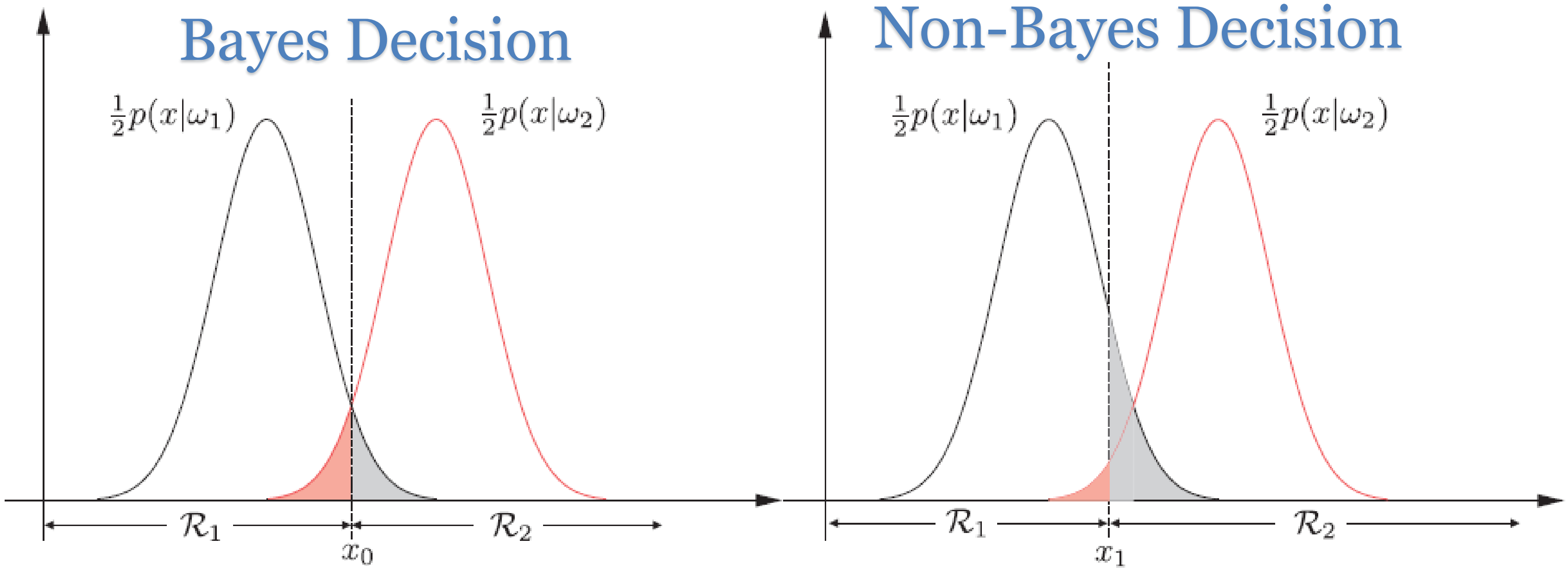
$$P(\omega_j|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_j)P(\omega_j)}{p(\boldsymbol{x})}, \quad j = 1, 2, \ldots, M,$$

$$\text{Assign } \boldsymbol{x} \text{ to } \omega_i = \arg\max_{\omega_j} p(\boldsymbol{x}|\omega_j)P(\omega_j), \quad j = 1, 2 \ldots, M.$$

Supervised Learning Models

› Bayes Classifier (Optimal)

– Illustration for equal probability cases (1D feature, M=2, $P(\omega_1)=P(\omega_2)=0.5$)



$$P_e = P(\omega_2)\int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2)d\mathbf{x} + P(\omega_1)\int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1)d\mathbf{x} : \quad \text{Probability of Error.}$$

Supervised Learning Models

- Bayes Classifier (Optimal):
  - Perquisites:
    - Conditional Probability: $P(x|\omega_k)$
      - Pdf estimation methods (Maximum Likelihood, Parzen windows, ….)
    - Prior Probability $P(\omega_k)$
      - Population Based: $P(\omega_k) \approx \frac{N_k}{N}$

Supervised Learning Models

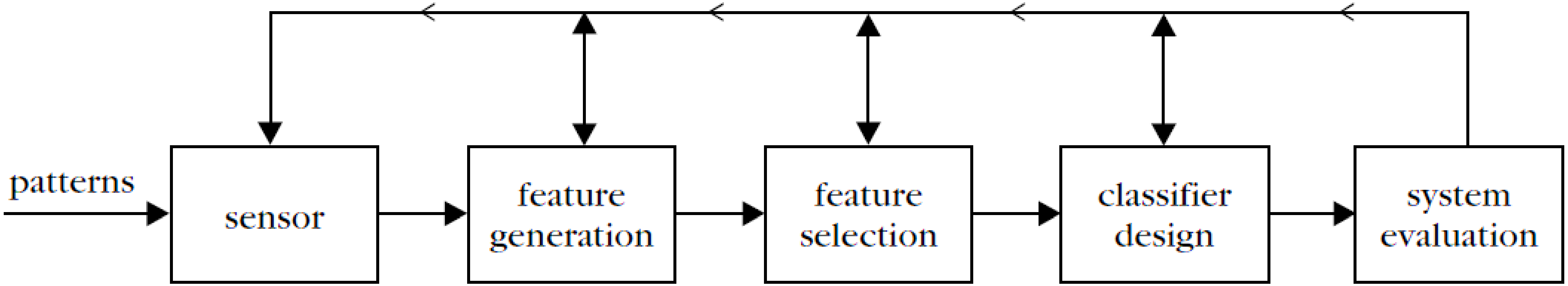
- Others:
  - Bayesian Network,
  - Meta Heuristic (Genetic Algorithm, ….)
  - …

Machine Learning Applications

- Computer Vision:
  - Object Detection/Recognition/....
- Authentication:
  - Face/Voice/Finger-Print/Iris/Retina/Gesture/Handwriting/ ..... Recognition
    - Identification
    - Verification
- Natural Language Processing/Understanding
- Speech Processing
- Search Engine/Social Media/....
- CAD (Computer Aided Diagnosis)

Supervised Learning Models

› Classification System Anatomy (Classical but Still alive!)



$x \in \Re^D$

$\otimes$

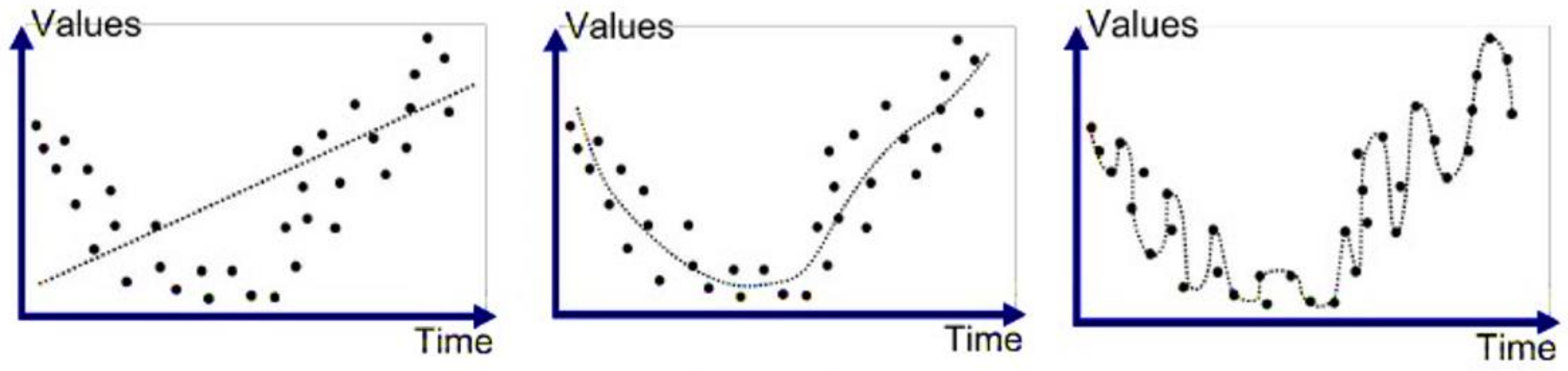
$x \in \Re^d$

$d \ll D$

$\otimes$

Supervised Learning Models

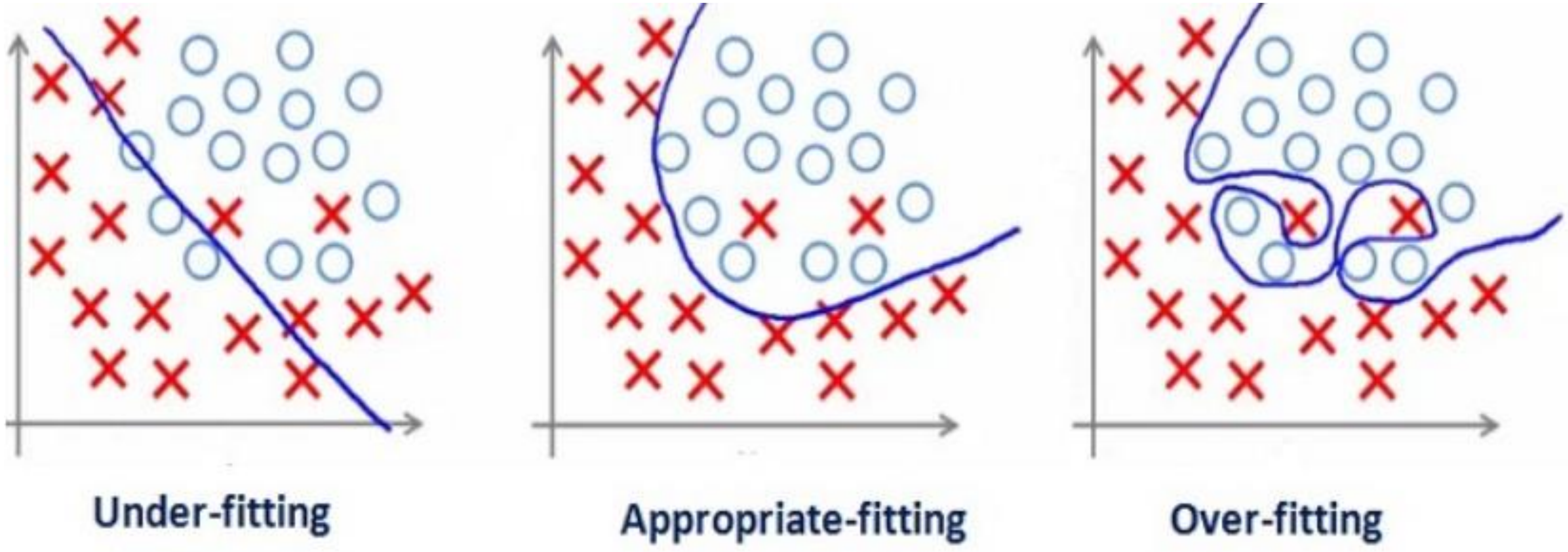
- Model Fitness:
- Two major problem: Overfitting/Underfitting
- Regression:



Underfitted   Good Fit/Robust   Overfitted

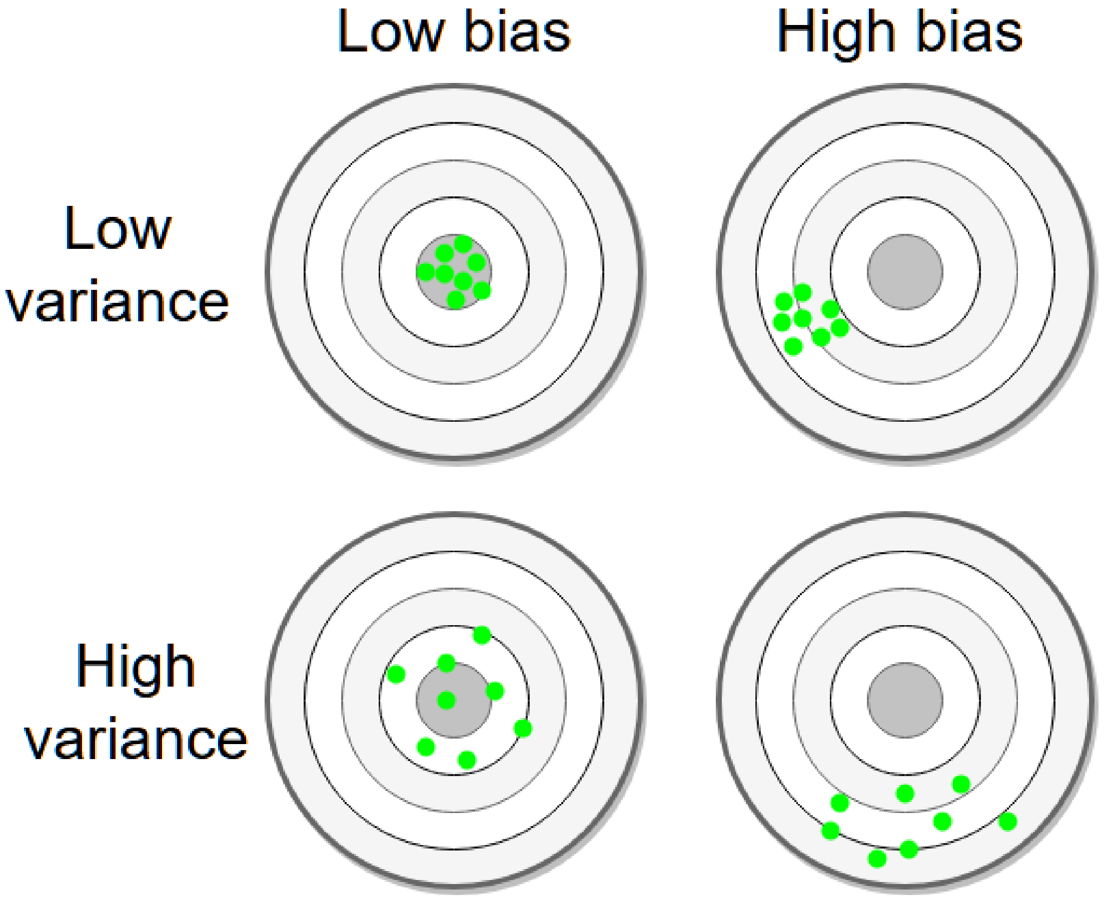
Supervised Learning Models

- Model Fitness:
- Two major problem: Overfitting/Underfitting
- Classification:



Supervised Learning Models

- Bias-Variance Dilemma:
  - Center: True Model
- Too Simple Model:
  - High Bias-Low Variance
- Too Complex Model:
  - Low Bias-High Variance



Supervised Learning Models

- Suppose:

$$y = f(x) + \varepsilon, \quad \{x_i, y_i\}_{i=1}^{N}, \quad \varepsilon \sim N\left(0, \sigma_e^2\right)$$

- We estimate a model:

$$y = \hat{f}(x)$$

- Estimation Error:

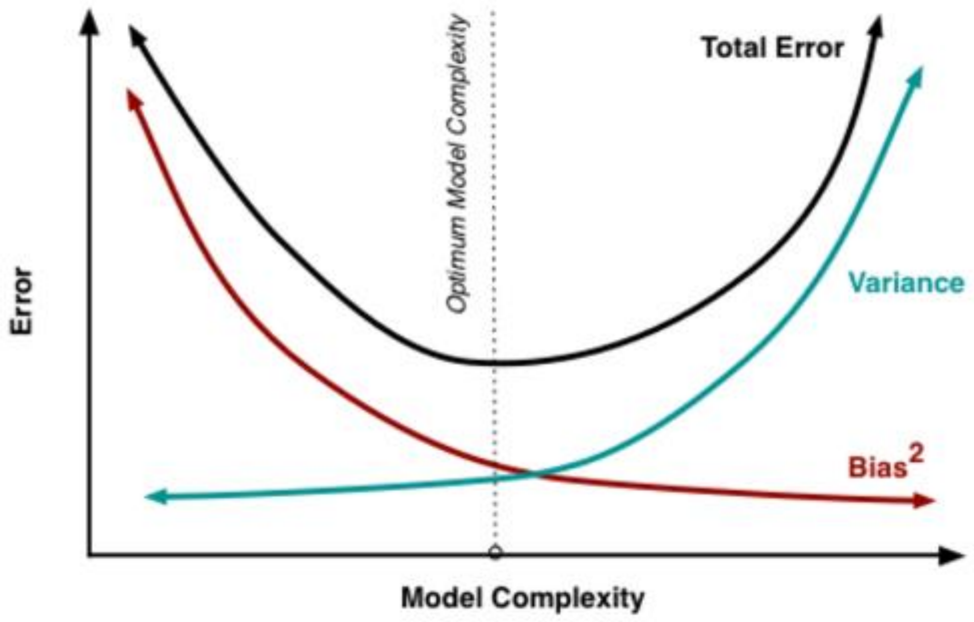
$$Err(x) = E\left|(Y - \hat{f}(x))^2\right|$$

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$

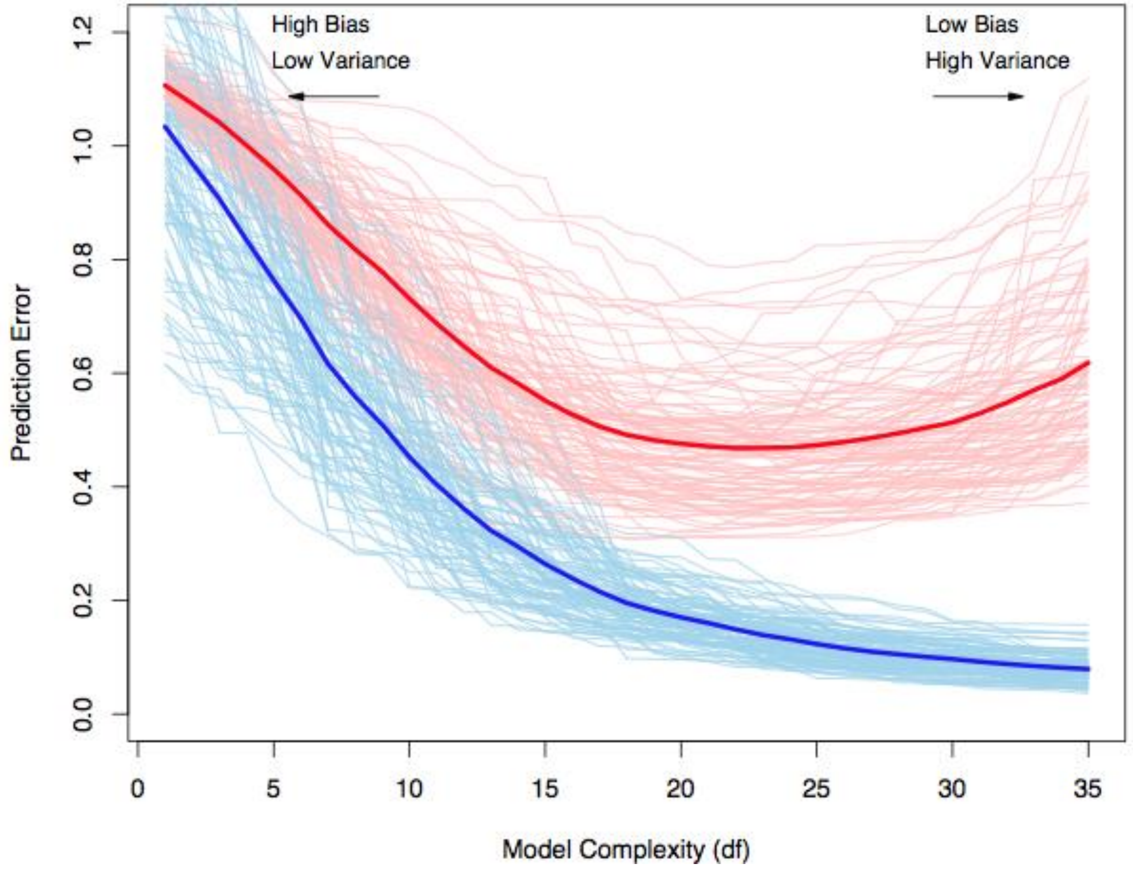
Supervised Learning Models

› Bias-Variance Dilemma:



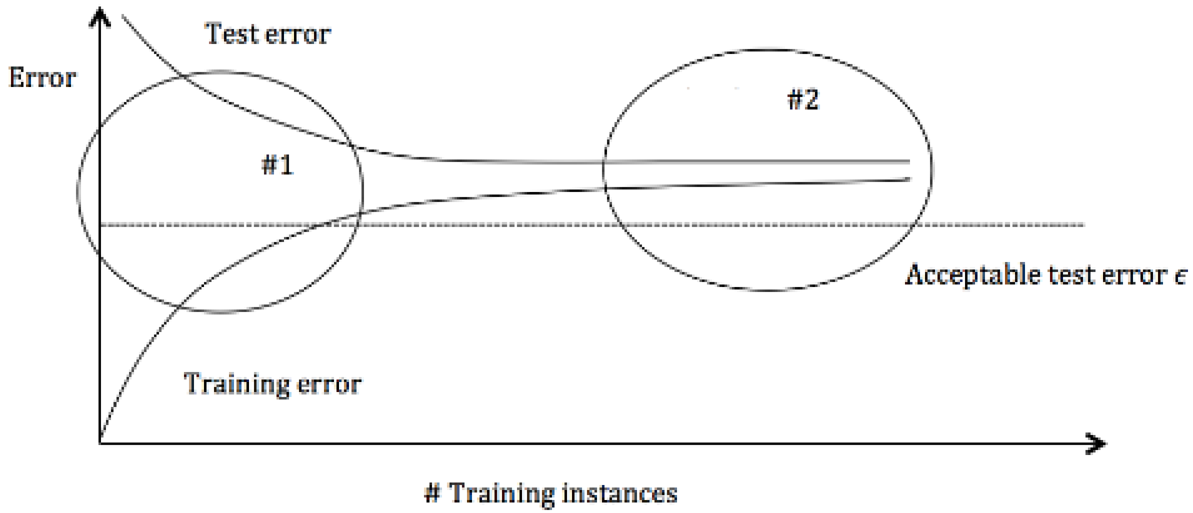
Supervised Learning Models

› Bias-Variance Dilemma:
  - 100 Run with 50 samples



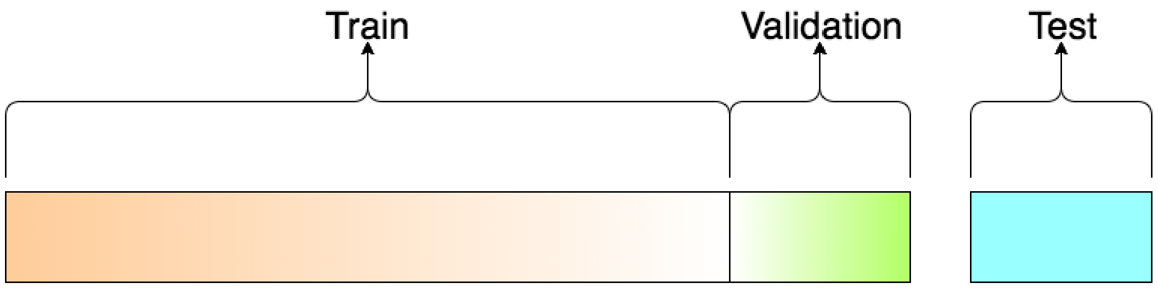
Supervised Learning Models

- Bias-Variance Dilemma:
  - #1: High Variance
  - #2: High Bias



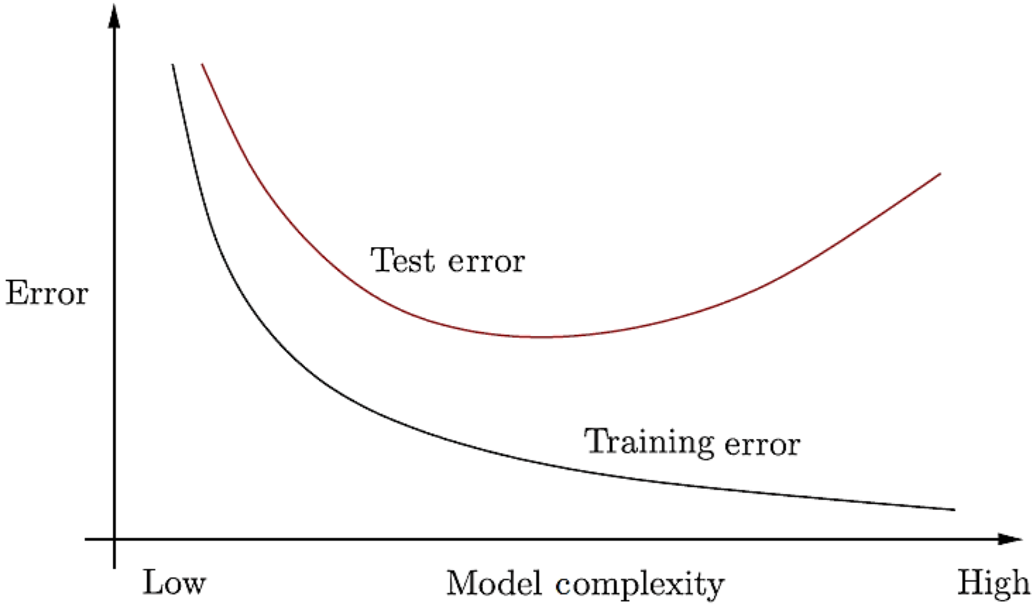
Machine Learning Evaluation

- How to Use Labeled Data?
  - Training Set:
    - Fit/Train the model
  - Validation Set:
    - Estimate Error for Model Selection (Hyperparameter Selection: # of Layer)
    - Model Sees this data, but never learn/tune with it
  - Test Set:
    - Generalization Error of ***finally*** Selected Model



Machine Learning Evaluation

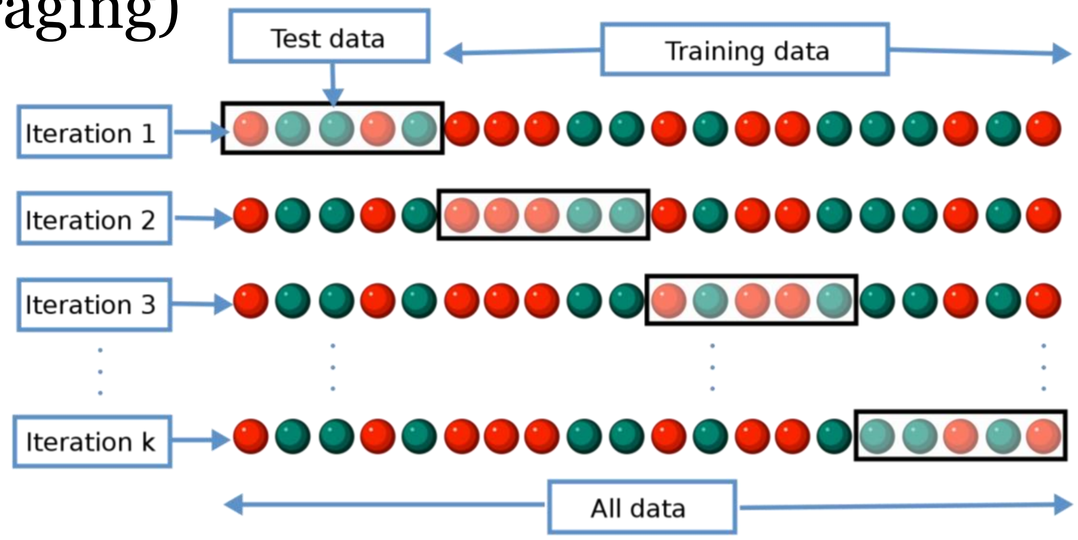
› Typical Performance:



K-Fold Cross Validation

› Goal: More Accurate Estimate of Model Prediction Performance

1. Data set is partitioned into, **K** (Typically 10), roughly equal-sized parts
2. Perform training **K** times
3. In each training: One Part for test, (**K**-1) Part for Training
4. Combine results (averaging)

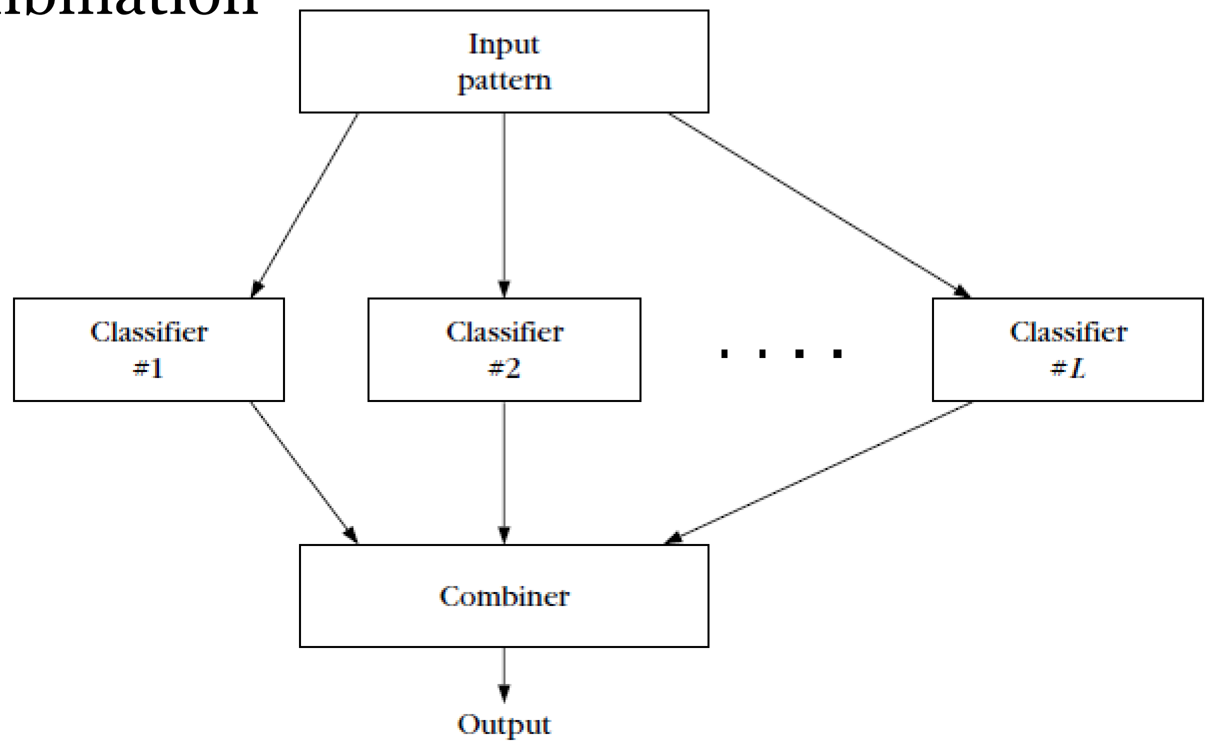


K-Fold Cross Validation

- **K**=N ➔ Leave-one-Out (LOO) Cross Validation
- Bootstrap:
  - Drawing random sample of size (~N/**K**), N times!

How to Enrich Classifier:

› Classifier Combination:
  - Majority Votes
  - Mathematical combination
  - Bagging
  - Boosting
  - ....



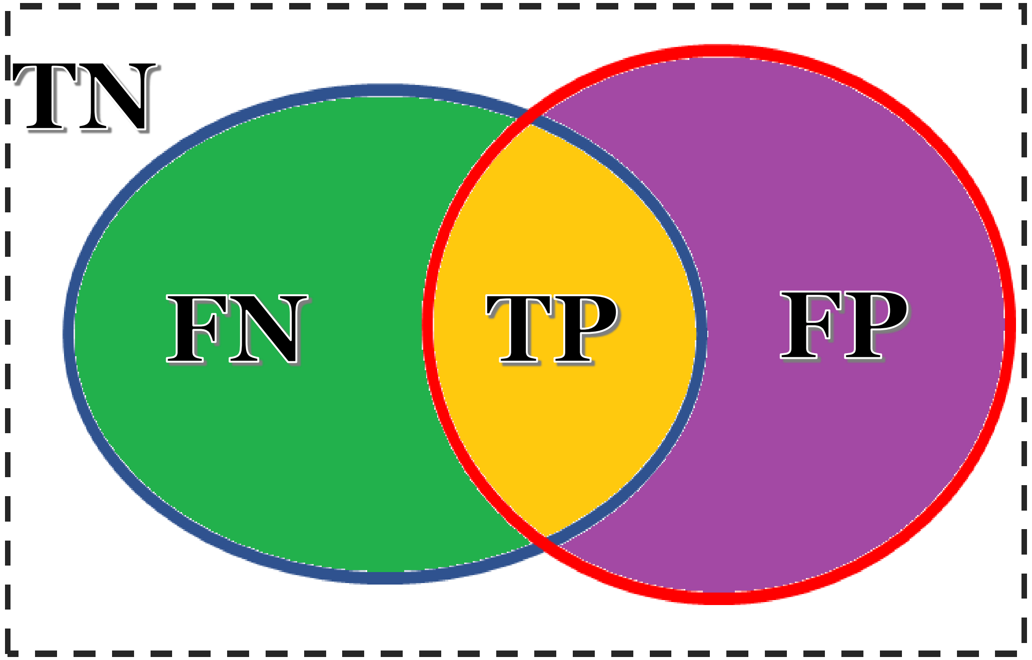
Machine Learning Evaluation

› Confusion Matrix:
  - Consider two-class binary classification
  - True/False-Positive/Negative Decision

		True Label	
Predicted Label		**Positive**	**Negative**
	Positive	**TP**	**FP**
	Negative	**FN**	**TN**

Machine Learning Evaluation (Repeated)

- Sensitivity, recall: $\dfrac{TP}{TP+FN}$
- Specificity: $\dfrac{TN}{TN+FP}$
- Precision: $\dfrac{TP}{TP+FP}$
- Accuracy: $\dfrac{TP+TN}{TP+TN+FP+FN}$
- Dice: $\dfrac{2TP}{2TP+FP+FN}$

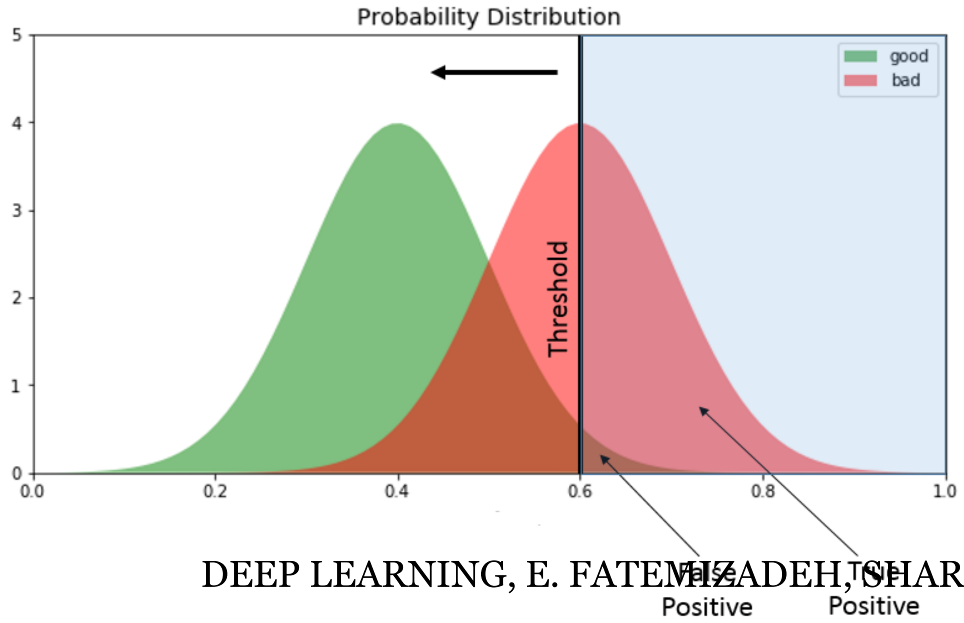


Machine Learning Evaluation

› ROC and AUC

› ROC (Receiver Operating Characteristic)

Performance Measurement for Two-Class Classification at Various Thresholds (One Dimensional Feature)

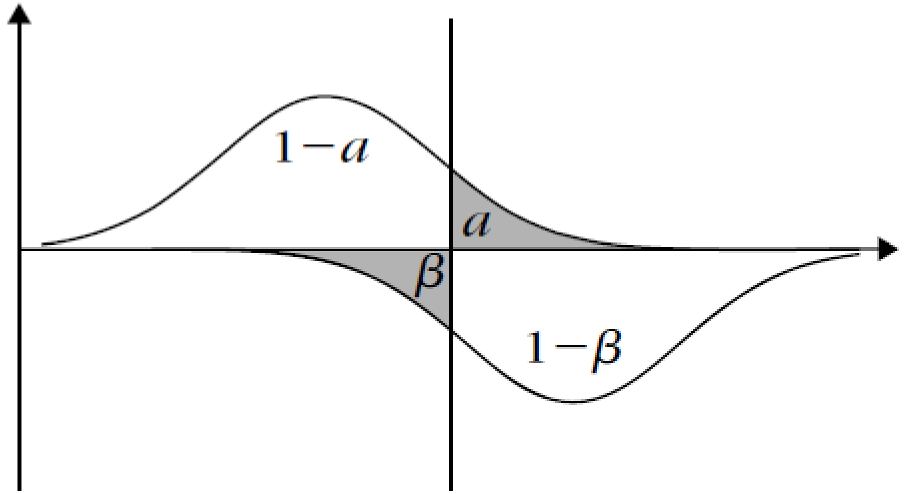


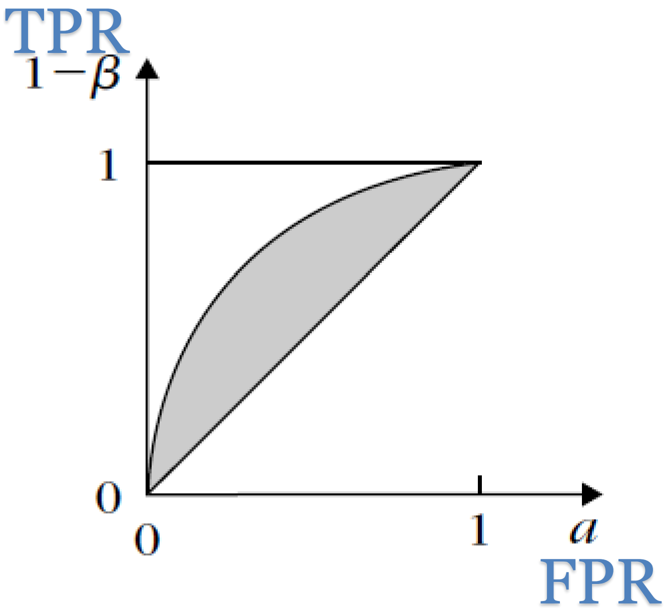
$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$$

Machine Learning Evaluation

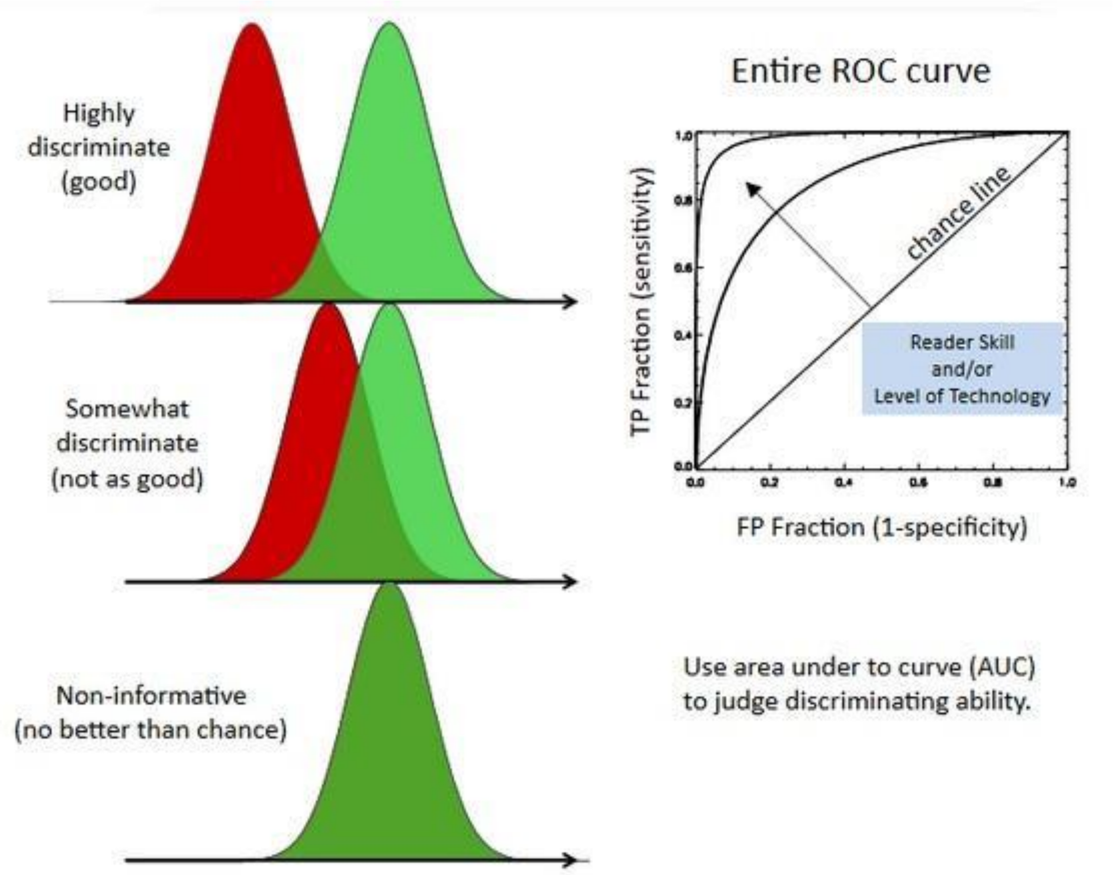
› ROC is TPR against FPR





› AUC: Area Under Curve

Machine Learning Evaluation



Machine Learning

› Question?