Foundations of Machine Learning
AI2000 and AI5000

FoML-17
Perceptron

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence
& Learning Lab

So far in FoML

- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions (regularization, model selection)
    b. Bias-Variance Decomposition (Bayesian Regression)
    c. Decision Theory - three broad classification strategies
        - Probabilistic Generative Models - Continuous & discrete data
        - (Linear) Discriminant Functions - least squares solution

The Perceptron

The Perceptron Algorithm

- Input: $x \in \mathbb{R}^D$
- Targets (2 classes): $t \in \{C_1, C_2\}$
- Prediction: $y(\mathbf{x}) = f(\mathbf{w}^T \phi(\mathbf{x}))$

Activation function f(a)

The Perceptron Algorithm

- Class decisions:
  - Assign x to C_1 if:
  - Assign x to C_{-1} if:

- Criterion for correct classification:

The Perceptron Algorithm

- The loss (perceptron criterion):

$$E_P(\mathbf{w}) =$$

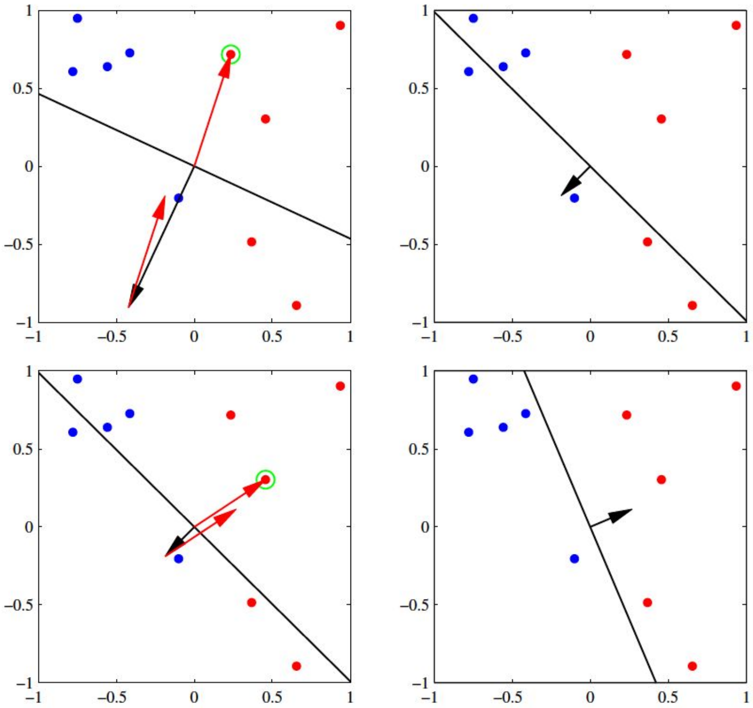
Perceptron learning: SGD

$$E_P(\mathbf{w}) = \sum_{n \in \mathcal{M}} \mathbf{w}^T \phi(\mathbf{x}_n) t_n$$
$$= \sum_{n \in \mathcal{M}} E_n(\mathbf{w})$$

SGD: for each misclassified sample x_n:

$$\mathbf{w}^{t+1} = \mathbf{w}^t -$$

Perceptron learning: SGD



If data is linearly separable, perceptron converges

Perceptron - Issues

- Works only for 2 classes
- More than one solutions
  - Initialization and the order in which the data is presented
- Will not converge if the dataset is not linearly separable
- Need to define basis functions
  - This is the case for all the methods that we discussed so far