Midterm Review

10/16/2023

Part IA: Loss Fn Based Approaches

$L(y, \hat{y})$, $\hat{y} = h_\theta(x)$

↓

$L(y, h_\theta(x))$

Training.

$D = \{(x^{(1)}, y^{(1)}), \ldots (x^{(m)}, y^{(m)})\}$

$$\min_\theta \sum_{i=1}^{M} L(y^{(i)}, h_\theta(x^{(i)}))$$

(1) Regression: $L(y, \hat{y}) = [y - \hat{y}]^2$

(2) Classifn: $L(y, \hat{y}) = 1(y \neq \text{sign}(\hat{y}))$

Problem: Cannot optimize. Not Continuous

Perceptron.

$L(y, \hat{y}) = \max(0, -y \cdot \hat{y})$

SVM

$L(y, \hat{y}) = \max(0, 1 - y \cdot \hat{y})$

optimize using GD

$\theta^{(0)} =$ Something.

for $t = 1 : T$

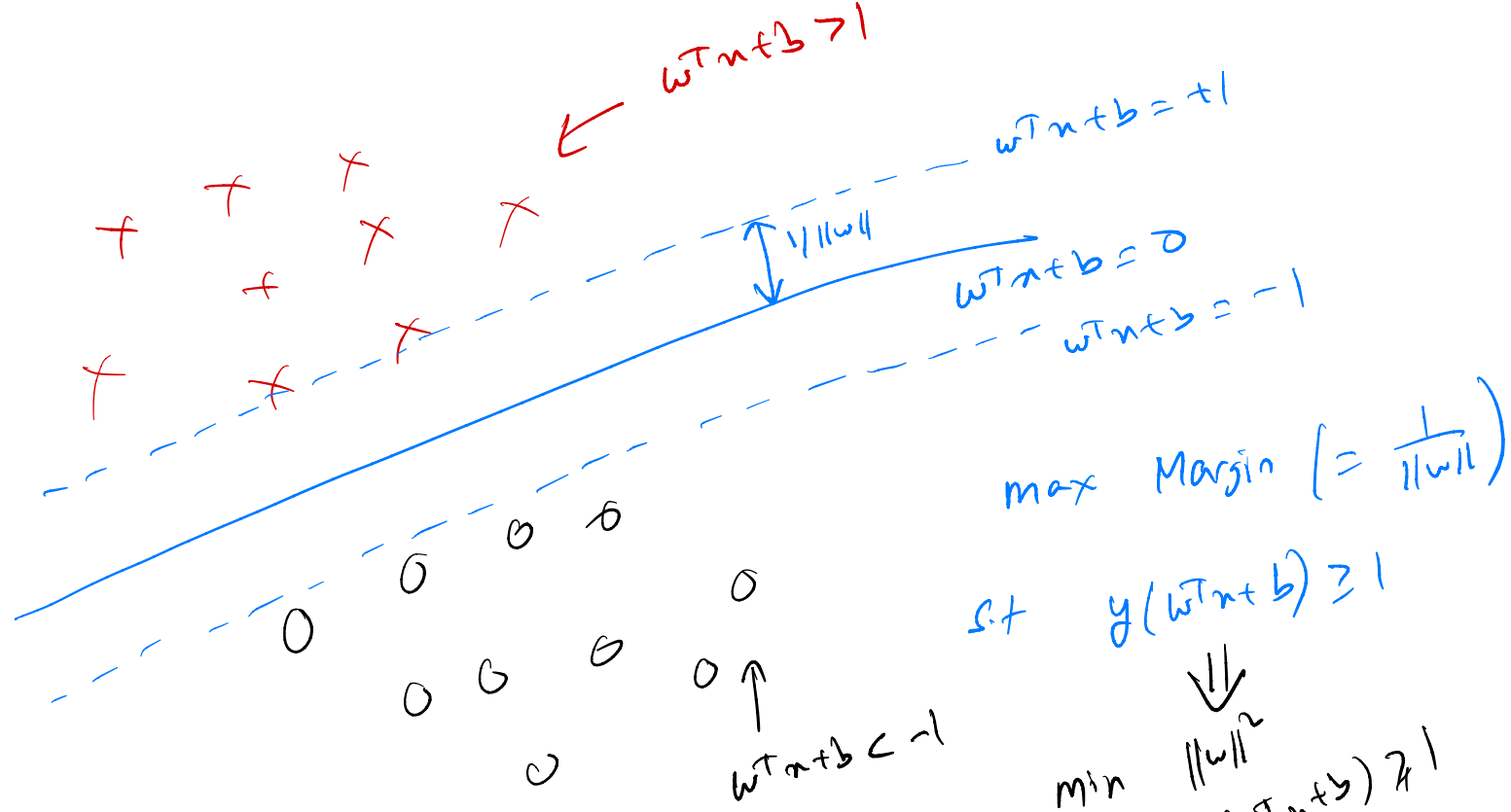
$\theta^{(t+1)} = \theta^t - \alpha \nabla L(\theta^t)$

Regularization

$$\min_\theta L(\theta) + \frac{\lambda}{2}\|\theta\|^2$$

Low $\|\theta\| \rightarrow$ Reduce overfit

Part II: Support Vector Machines



max Margin $\left(= \frac{1}{\|w\|}\right)$

s.t $y(w^T x + b) \geq 1$

$\Downarrow$

min $\|w\|^2$

s.t $y(w^T x + b) \geq 1$

Part II: Non-Parametric Approaches

KNN

No training
(Store Data in Mem)

Test: $\hat{y} = y^{(i^*)}$

where $i^* = \underset{i}{\operatorname{argmin}} \|x^{(i)} - x\|^2$

- Normalization important
- Careful with irrelevant attributes
- High inference time/cost, High memory

→ kD trees to reduce Inf. time

Decision Trees

Loss fn: $\min_T \sum L(h_T(x), y)$

DT Algo: Greedy Algo

Achieve Certainty or instances of "same Class" on the same side after splitting

- Entropy

$\arg\max_{i,t} H(Y) - H(Y \mid [x_i \leq t])$

Part III: Probabilistic Models

$p(y \mid x, \theta)$, $p(x, y \mid \theta)$

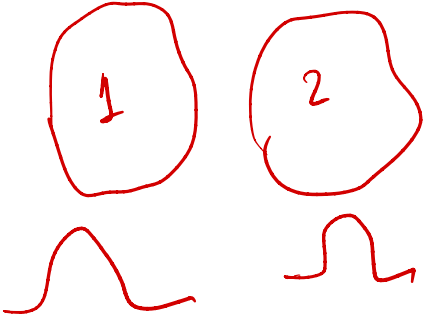
↑ Conditional. ↑ Joint.

$\sigma(y, x, w, b) = \frac{1}{1+\exp(-y(w^T x + b))}$

↓

Generative Models (NB)

Model $p(x, y \mid \theta) = p(x \mid y, \theta)\, p(y \mid \theta)$



Learning: Max Likelihood

NB: $p(x \mid y) = \prod_{i=1}^{d} p(x_i \mid y)$

Inference:

$\underset{y}{\operatorname{argmax}}\ p(y \mid x)$

Discriminative Models (LR)

Model: $p(y \mid x, \theta)$

Ideal: $p(y = \hat{y} \mid x, \theta) = 1(y \text{ is } \hat{y} \text{ else } 0)$

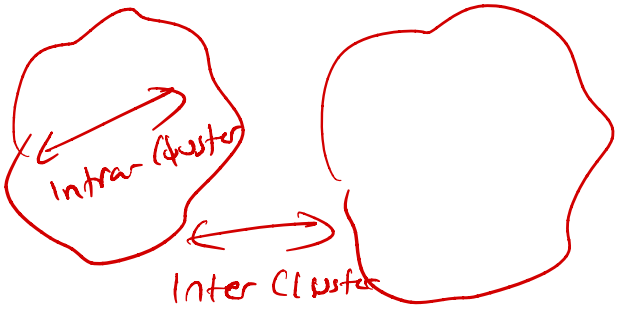
Since above is non-continuous

$p(y \mid x, \theta) = \sigma(y, x, \theta)$

Learning: Cond Log Likelihood

Inference: $\underset{y}{\operatorname{argmax}}\ p(y \mid x, \theta)$

Part IV : Unsupervised Learning



K-Means

- Assignment step (AS)
- Means step (MS)

Alternating Minimizⁿ

$$\min_{S_1 \dots S_K,\, M_1 \dots M_K} \sum_{i=1}^{K} \sum_{j \in S_i} \|x_j - M_i\|^2$$

AS : Fix M, optimize S

MS : Fix S, optimize M.