Midterm Review

03|28|2023

Part I: Loss Fn Based Approaches

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

$\Downarrow$

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

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

② Classif^n: $L(y,\hat{y}) = \mathbb{1}(y \neq \hat{y})$

Problem: Cannot optimize.

Perceptron.

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

SVM

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

Training.

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

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

$\uparrow$

$\underbrace{\text{optimize using GD}}$

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

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 $\|\theta\| \to$ Reduce overfit.

Part II: Non-Parametric Approaches

kNN

No training
(Store Data in Mem)

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

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

- Normaliz^n important
- Careful with irrelevant attributes
- High inference time / Cost, High memory

→ kD trees to reduce Inf. time

Decision Trees

$x_i \le t$

Y ↙ ↘ N

Loss fn: $\min_T \sum_i L(h_T(x^i), y^i)$

DT Algo: Greedy Algo

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

– Entropy

$$\underset{i,t}{\arg\max}\ H(Y) - H(Y \mid \{x_i \le t\})$$

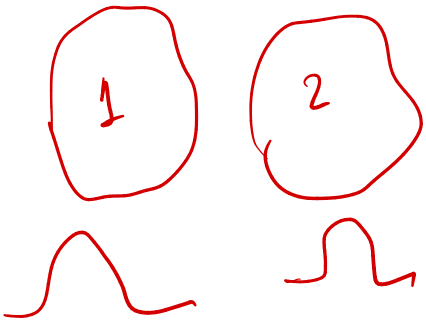
Part III: Probabilistic Models

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

$\uparrow$ Conditional. $\uparrow$ Joint.

Generative Models (NB)

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



Learning:

Max Liklihood.

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

Inference:

$\arg\max_{y} p(y \mid x)$

Discriminative Models (LR)

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

Ideal: $p(y = \hat{y} \mid x, \theta) = 1(y \ \& \ \hat{y}$ are same$)$

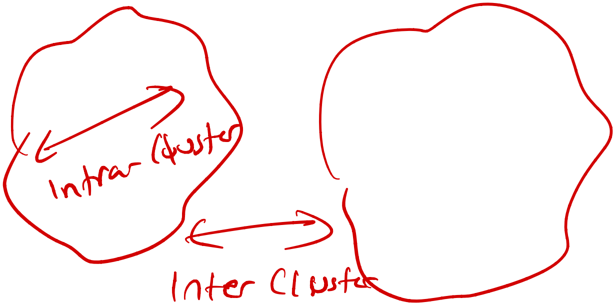
Since above is non-diff,

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

Learning: Cond Log Liklihood

Inference: $\arg\max_{y} p(y \mid x, \theta)$

Part IV: Unsupervised Learning



K-Means

— Assignment Step (AS)

— Means Step (MS)

Alternating Minimiz^n

$$\min_{S_1 \dots S_K,\ \mu_1 \dots \mu_K} \sum_{i=1}^{K} \sum_{j \in S_i} \|x_j - \mu_i\|^2$$

AS: Fix μ, optimize S

MS: Fix S, optimize μ.