COL774
Machine Learning
Oct 20, 2021

Unsupervised Learning :-

Till Now:- Supervised setting

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$:- Training data

$h_\theta(x) = ?$ $\quad x^{(i)} \in R^n \;|\; x^{(i)} \in \{--\}$ $\{0,1\}$

classification $\begin{cases} y^{(i)} \in \{0,1\} \\ y^{(i)} \in \{0, 1 --- r\} \end{cases}$

Regression $\{ y^{(i)} \in R$

Linear Reg., Logistic Reg.. - - - Neural Networks, Deep Networks

Unsupervised Learning $\{x^{(i)}, \cancel{y^{(i)}}\}_{i=1}^{m}$ $\quad$ Data:- $\{x^{(i)}\}_{i=1}^{m}$ $\quad x^{(i)} \in R^n$

"Learning" | Pattern finding.

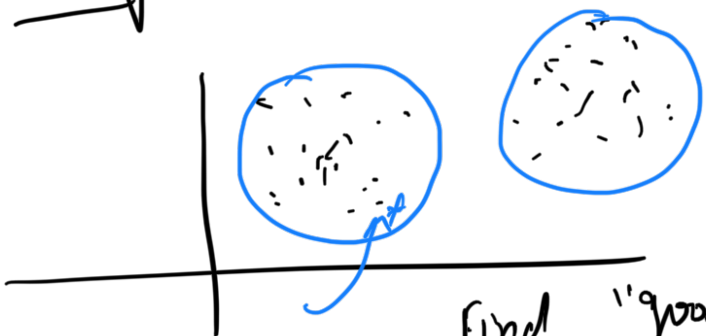
$\{x^{(i)}\}_{i=1}^{m}$

$\{x^{(i)}, z^{(i)}\}_{i=1}^{m}$
↳ hidden
$p(x^{(i)}, z^{(i)}; \theta)$

semi-supervised setting $\{x^{(i)}, \cancel{y^{(i)}}\}_{i=1}^{m}$

A subset of $y^{(i)}$'s are available ↳ $\begin{cases} \text{some are available} \\ \text{others are not} \end{cases}$

① Clustering :- $\{x^{(i)}\}_{i=1}^{m}$ $\quad x^{(i)} \in R^n$



K = 2
K-Means

Find "good" clusters
↳ what is good?

② Density Estimation :-

↳ GMM (Gaussian Mixture Models) ← GDA

↳ EM (Expectation Maximization)

↳ To deal with missing data

→ $\underbrace{p}(x^{(i)};\theta)$]

↳ "outliers"

Model:- $\underline{p(x^{(i)}, z^{(i)};\theta)}$

Joint Distribution

$$\underbrace{p(x^{(i)};\theta)} = \sum_{z^{(i)}} \underbrace{p(x^{(i)}, z^{(i)};\theta)}$$

$$p(x^{(i)};\theta) = \sum_{z^{(i)}} p(x^{(i)}, z^{(i)};\theta) = \sum_{z^{(i)}} \underbrace{p(x^{(i)}|z^{(i)};\theta)}\, p(z^{(i)};\theta)$$

↳ Gaussian :-

GMM:- Gaussian Mixture Model

→ "iterative algorithm"

$z^{(i)}$'s & θ

⇒ EM:- Expectation Maximization

$$\underline{p(x^{(i)}, z^{(i)};\theta)} \qquad \{x^{(i)}\}_{i=1}^{m}$$

$$\log\left[\prod_{i=1}^{m} p(x^{(i)};\theta)\right] = \log \prod_{i=1}^{m}\left[\sum_{z^{(i)}} p(x^{(i)}, z^{(i)};\theta)\right]$$

$$= \sum_{i=1}^{m} \log \underbrace{\sum_{z^{(i)}} p(x^{(i)}, z^{(i)};\theta)}$$

$$\geq \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}$$

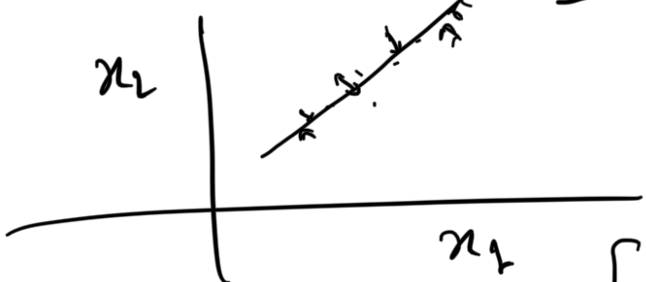
Jensen's inequality ←

E:- Expecation
M:- Maximization

Estimates of distributions our hidden variables

PCA:- Principal Component Analysis

$x^{(i)} \in \mathbb{R}^n$:- Project the data onto a lower dimensional subspace.



$1 \,--\, K$

① Compact Rep (Noise free)

② :- Smaller Features

$\leftarrow [\{u_1 \,--\, u_K\}$:- Principal Components

$$x^{(i)} = \sum_{k=1}^{K} z_k^{(i)} u_k \qquad u_k \in \mathbb{R}^n$$

FINALLY:-

Learning Theory.

Theoretical Foundations of Machine Learning

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$:- $(h_\theta(x))$

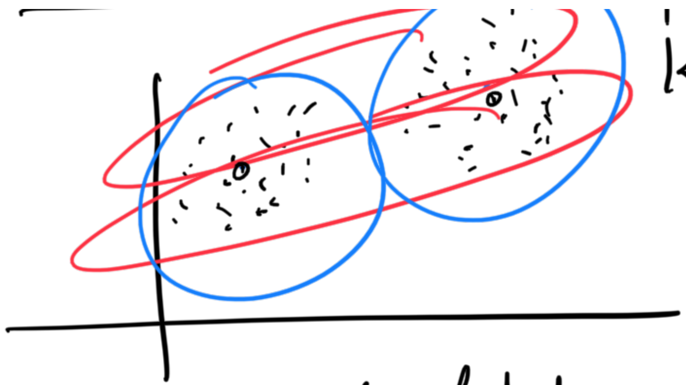
↳ Generalize well,

$$\Pr(|\hat{\varepsilon}(h) - \varepsilon(h)| \geq \delta) \leq 2e^{-2\gamma^2 m}$$

Training error — Error in unseen data — $0 \leq \delta \leq 1$ — Parameter

PAC:- Probably Approximately Correct

K- Means Clustering :- # of clusters Given:-



K $\{x^{(i)}\}_{i=1}^{m}$ $x^{(i)} \in R^n$

Find :-

(I) $\mu_1 \ldots \mu_K$:- cluster centroids

$\mu_k \in R^n$

(II) $\forall i$:- $c^{(i)} \in \{1 \ldots K\}$

↳ cluster assignment for $x^{(i)}$

"good" clusters.

Notion of a good clustering :-

(a) Intra cluster distance is minimized

(b) Inter cluster distance will be maximized

$$\underset{\substack{\{\mu_k\}_{k=1}^{K} \\ \{c^{(i)}\}_{i=1}^{m}}}{\arg\min} \sum_{i=1}^{m} \|x^{(i)} - \mu_{c^{(i)}}\|^2$$ is minimized

$\{\mu_1 \ldots \mu_K\}$

$\{c^{(1)} \ldots c^{(m)}\}$

Mixed optimization problem.
(Discrete + Cont.)

K-Means Algorithm :-
(Iteration)

(A) If μ_k's are known :-

$\forall i$:- $c^{(i)} = \underset{K}{\arg\min} \|x^{(i)} - \mu_k\|^2$

(B) If $c^{(i)}$'s are known :-

$$\mu_k = \frac{\sum_{i=1}^{m} 1\{c^{(i)}=k\}\, x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)}=k\}}$$

$$\mu_k = \underset{\mu_k}{\arg\min} \sum_{i=1}^{m} 1\{c^{(i)} = k\} \| x^{(i)} - \mu_k \|^2$$ ✓