Unsupervised Machine learning

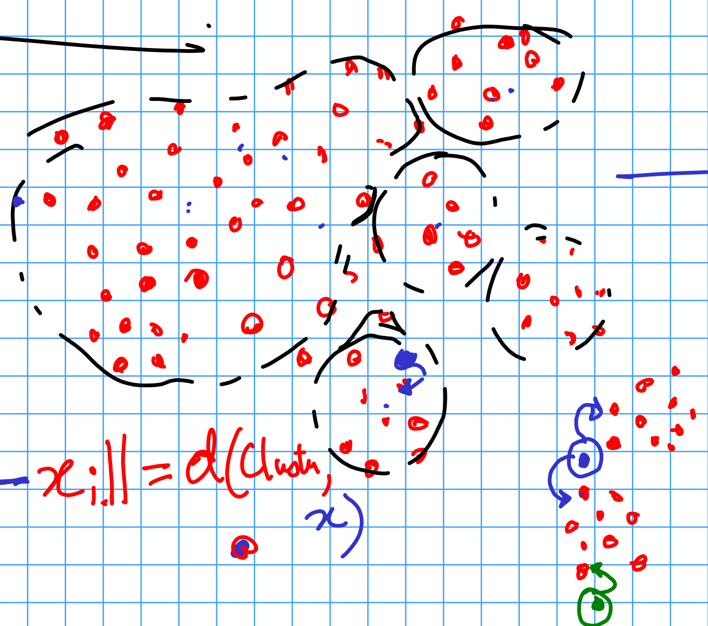
Two ways → Crowd sourcing
→ Unsupervised ML

Until now, f: the predictor is given direct supervision.

$(x, y) \rightarrow \boxed{\text{Regression/ classifier}} \rightarrow f(x)$

Suppose we ~~have~~ know labels of very few data. Given a new data w/o labels, you simply ~~...~~ check which existing data is close to the new data.

(*)



→ ~~Nearest~~ Nearest-Nbr Classifier / Clustering

$$\frac{1}{|Cluster|}\sum_{i \in Cluster} \|x - x_i\| = d(Cluster, x)$$

$\|x - x_i\| \sim m + \epsilon$

* Given a set of features $\{x_i\}_{i \in [N]}$ ($1 \ldots N$)
how to cluster these points?

REQD: (1) Notion of distance: Euclidean distance
OR Notion of similarity: Similarity $\simeq$ Inverse measure of distance

(2) Number of clusters K

* Given a set of features $\{x_i\}_{i \in [N]}$ how to cluster these points?

REQD: ① Notion of distance ; Euclidean distance
or Notion of similarity ; Similarity $\approx$ Inverse measure of distance

② Number of clusters K

Formal statement?

Output

Input : $\{x_i\}_{i \in [N]}$ → ① Cluster label c for each x_i

② Centroid of cluster μ_c for each c

Given ① distance (x_i, x_j)
$= \|x_i - x_j\|$

② # Cluster $= K$
Clusters are indexed as $1 \ldots K$

$x_i \to c \to K$
$\downarrow$
N

$P_{ic} \to i$ belongs to cluster c

* Opt. variable for ①
$P_{ic} = 1$ if i belongs to cluster c
$= 0$ o/w.

* Opt. variable is μ_c.

$\{x_i\}_{i \in 1 \ldots N} = \{x_1 \ldots x_n\}$
$[K] \ldots = \{1, 2, \ldots K\}$
$\Rightarrow \begin{bmatrix} P_{ic} \\ \mu_c \end{bmatrix} \to$ output $\forall i, c$.

Probabilistic
P_{ic} = the prob. that i belongs to c

$\sum_{c=1}^{K} P_{ic} = 1$

Deterministic.
$P_{ic} = 1$ if $i \to c$
$= 0$ o/w

$x_1 \ldots \ldots x_n$ $\quad$ P_{ic} and μ_c $\forall\, i, c$.

$1 \ldots \ldots K$

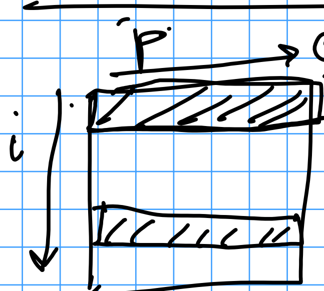
objective: $$\underbrace{\min_{P_{ic}, \mu_c} \sum_{i=1}^{N} \sum_{c=1}^{K} P_{ic} \|x_i - \mu_c\|^2}$$

$$\sum_{c=1}^{K} P_{ic} = 1.$$

Fix i: $P_{ic} = 1$ for some c

$= 0$ for others

Say the optimal mean μ_c^* for cluster c is



$\sum_{c=1} P_{ic} = 1.$

$$\min_{P_{ic}} \sum_{i=1}^{N} \sum_{c=1}^{K} P_{ic} \| x_i - \mu_c^* \|^2 = F(P, \underline{\mu_c^*})_{c=1..K}$$

$$= \sum_{i=1}^{N} \min_{P_{ic} :\ \sum_{c=1}^{K} P_{ic} = 1} \sum_{c=1}^{K} P_{ic} \| x_i - \mu_c^* \|^2$$

$$\min_{P_{ic}} \sum_{c=1}^{K} P_{ic} \underbrace{\| x_i - \mu_c^* \|^2}$$

so that $\sum_{c=1}^{K} P_{ic} = 1$

$$P_{ic} = 1 \quad \text{when } c = \underset{c \in [k]}{\operatorname{argmin}} \|x_i - \mu_c^*\|$$
$$= 0 \quad \text{o/w}$$

$$\min_{\mu_c^*} \; \min_{P} \; F(P, \mu_c^* \mid c = 1 \dots k)$$

follow these iterations

$$P^t \leftarrow \underset{P}{\operatorname{argmin}} F(P, \mu_c^{*(t)} \mid c = 1 \dots k)$$

$$\mu_c^{*(t+1)} = \underset{\mu_c}{\operatorname{argmin}} F(P^t, \mu_c \mid c = 1 \dots k)$$

Gradient Descent