Model Selection problem.

Eg. 1. $m(x) = E(Y \mid X = x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p.$

$(M_1, M_2, \cdots, M_p)$ $\quad$ M_k is the kth degree polynomial.

Eg 2. Age of Death ~~rate~~ certain population. $(Y_1 Y_2 \cdots Y_n)$.

1. M_1 = exponential.
2. M_2 = gamma distn.
3. M_3 = lognormal.

Eg 3. AR model. $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} \cdots + a_k Y_{t-k} + \epsilon_t.$

the order of the model as well as the parameters are unknown.

Eg 4. Mixture Model: $f(\underset{\sim}{y}) = \sum_{j=1}^{k} \pi_j \, \phi(\underset{\sim}{y}, \underset{\sim}{\mu}_j, \Sigma_j)$

$\sum_{i=1}^{k} \pi_i = 1$ or $\pi_i < 1$ $\quad$ k - unknown. $\quad$ $\underset{\sim}{\pi}$ is unknown.

$\underset{\sim}{\mu}_j, \Sigma_j$ all are unkn.

③

$$N(\overset{5,\ 1}{\mu_1, \sigma_1^2}), \quad N(\overset{8,\ 1}{\mu_2, \sigma_2^2})$$

$\pi_1 = 0.3$ $\qquad$ $\pi_2 = 0.7$

$\pi_1 + \pi_2 = 1.$

$$T = \frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma} \qquad \frac{1}{n-1}\left(\Sigma(x_i - \bar{x})^2\right)$$

generate sample from mixture distribution. as specified above.

Correct. If $(U < \pi_1)$ generate from $N(\mu_1, \sigma_1^2)$

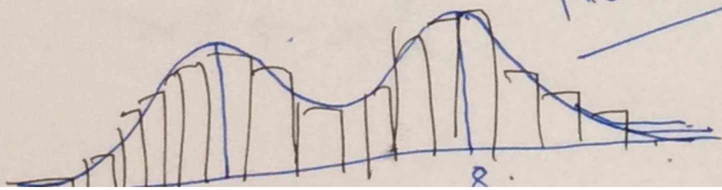
else. generate from $N(\mu_2, \sigma_2^2)$.

for each sample. Y_i ✓

for $i = 1, 2 \ldots n$.

$$\begin{cases} U_i \leftarrow U[0, 1] \\ \text{If } (U_i < \pi_1) \\ \quad Y_i \leftarrow N(\mu_1, \sigma_1^2) \\ \text{else} \\ \quad Y_i \leftarrow N(\mu_2, \sigma_2^2) \end{cases}$$

Not Normal.



Incorrect.

$X_1 \sim N(\mu_1, \sigma_1^2)$

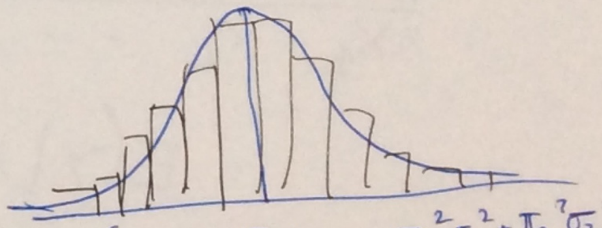
$X_2 \sim N(\mu_2, \sigma_2^2)$

$Y = \pi_1 X_1 + \pi_2 X_2$

Although the mean will be same.

For each iteration i. for $i = 1, 2, \ldots n$.

$$\begin{cases} X_{1i} \leftarrow N(\mu_1, \sigma_1^2) \\ X_{2i} \leftarrow N(\mu_2, \sigma_2^2) \\ Y_i = \pi_1 X_{1i} + \pi_2 X_{2i} \end{cases}$$



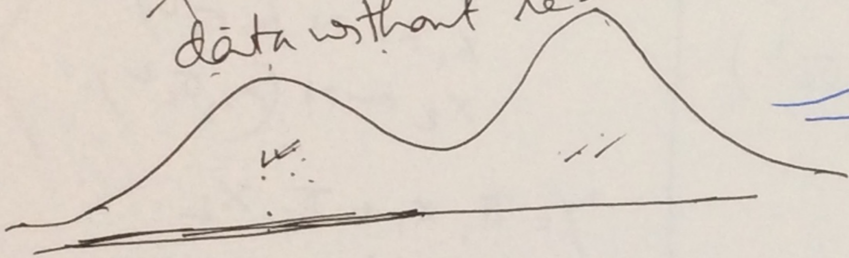
$N(\pi_1\mu_1 + \pi_2\mu_2, \ \pi_1^2\sigma_1^2 + \pi_2^2\sigma_2^2$

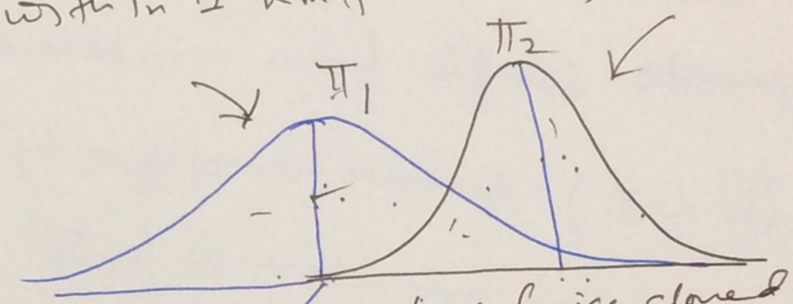
Distribution of home price (Y).

X = indicator function.

$$X = \begin{cases} 1 : & \text{if the home is with in 1 km from City,} \\ 0 . & \text{o/w} \end{cases}$$

data without label.





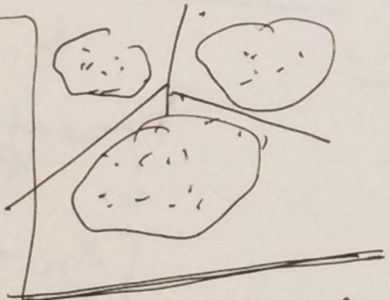
home price closed to city

home price away from city.

(data with label).

Normal family ✓
Mixture of $K=2$ population.

Estimate the parameter.



Un supervised.

a new data comes → make the labeling.

Classification

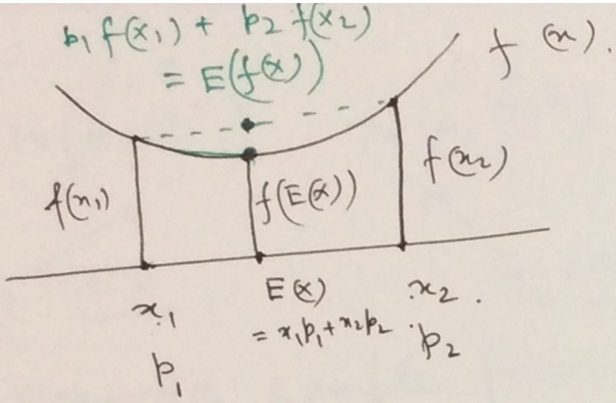
supervised.

EM algorithm to get the parameters.

$X_1 \quad X_2 \quad - - - - \quad X_n$

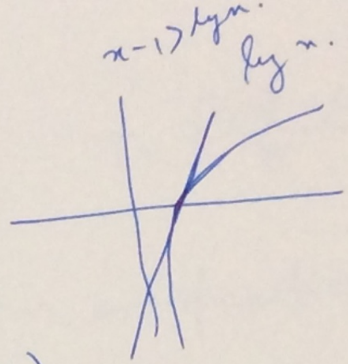
$X_i \sim$
$p_1 \; C_1 \to f_1$
$p_2 \; C_2 \to f_2$
$\vdots$
$p_k \; C_k \to f_k$

$E(Y \mid X=0)$



$p_1 + p_2 = 1$

$\lambda + (1-\lambda) = 1$



K-L divergence: To show $D_{KL}(P \| Q) \geq 0$

$$D(P\|Q) = \int p(x) \log_e \left(\frac{p(x)}{q(x)}\right) dx.$$

$$= -\int p(x) \log_e \left(\frac{q(x)}{p(x)}\right) dx.$$

$$\geq -\int p(x) \log \left(1 - \frac{q(x)}{p(x)}\right) dx.$$

$$= \int q(x)\, dx - \int p(x)\, dx = 0.$$