

Expectation-Maximization

■ Model Based Learning (Clustering)

Gaussian Model $N[\mu, \sigma^2]$

Classification Regress

■ Boot Step

■ - Initialize k clusters $C_1, \dots, C_k$

(μ_j, Σ_j) and $P[C_j]$ for each cluster j

K clusters

$\mu, \sigma, P[C] \leftarrow$ randomly assign

■ Iteration Step

■ - Estimate the cluster of each data $\Rightarrow$ Expectation

$P[C_j | x_i]$

■ - Re-estimate the cluster parameters $\Rightarrow$ Maximization

$(\mu_j, \Sigma_j), P[C_j]$ for each cluster j

$P[h|x] \leftarrow$ ND Model-based clustering

$P[\text{Hypothesis} | \text{Ovbservation Data}] = \text{Posterior} = P[h | D]$

$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$

$$P[y = i | x] = \frac{P[x | y = i] P[y = i]}{P[x]} = \frac{N[\mu_i, \Sigma_i] P[y = i]}{P[x]}$$

$$P[C_k | x_i] = \frac{P[x_i | C_k] P[C_k]}{P[x_i]} = \frac{N[x_i, \mu_k, \Sigma_k] P[C_k]}{P[x_i]} = \frac{P[x_i | C_k] P[C_k]}{\sum_k P[x_i | C_k] P[C_k]}$$

$$P[C_k] = \frac{\sum_i P[C_k | x_i]}{N}$$

$$P[x_i | C_k] = \frac{1}{\sqrt{2\pi} \sigma_k^2} e^{-\frac{(x - \mu_k)^2}{2\sigma_k^2}}$$

$$P[y = i | x] = \frac{\frac{1}{(2\pi)^{m/2} |\Sigma_i|^{1/2}} \text{Exp}[-\frac{1}{2} (x_k - \mu_i)^T \Sigma_i (x_k - \mu_i)] p_i}{P[x]}$$

Multivariate Normal Distribution

$$\frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2} (X-\mu)^\top \Sigma^{-1} (X-\mu)}$$

EM Classifier

	Classification Results
	$P[C_1 x_1]$
	...
	$P[C_j x_i]$
$x_1 = \{r_1, g_1, b_1\}$	
$x_2 = \{r_2, g_2, b_2\}$	
...	
$x_i = \{r_i, g_i, b_i\}$	
...	
	Cluster Parameters
	$(\mu_1, \Sigma_1), P[C_1]$ for C_1
	...
	$(\mu_k, \Sigma_k), P[C_k]$ for C_k

⇒ EM Classifier ⇒

■ Input (Know)

$x_1 = \{r_1, g_1, b_1\}$
 $x_2 = \{r_2, g_2, b_2\}$
 ...
 $x_i = \{r_i, g_i, b_i\}$
 ...

■ Output (Unknown)

■ Cluster Parameters

$(\mu_1, \Sigma_1), P[C_1]$ for C_1
 ...
 $(\mu_k, \Sigma_k), P[C_k]$ for C_k
 ...

■ Classification Results

$P[C_1 | x_1]$
 ...
 $P[C_j | x_i]$
 ...

Expectation Step

■ Input (Know)

$$\begin{aligned} \mathbf{x}_1 &= \{r_1, g_1, b_1\} \\ \mathbf{x}_2 &= \{r_2, g_2, b_2\} \\ &\dots \\ \mathbf{x}_i &= \{r_i, g_i, b_i\} \\ &\dots \end{aligned}$$

■ +

■ Input (Estimation)

■ Cluster Parameters

$$\begin{aligned} &(\mu_1, \Sigma_1), P[C_1] \text{ for } C_1 \\ &\dots \\ &(\mu_k, \Sigma_k), P[C_k] \text{ for } C_k \\ &\dots \end{aligned}$$

■ $\Rightarrow$

■ Output

■ Classification Results

$$\begin{aligned} &P[C_1 | \mathbf{x}_i] \\ &\dots \\ &P[C_j | \mathbf{x}_i] \\ &\dots \end{aligned}$$

$$P[C_j | \mathbf{x}_i] = \frac{P[\mathbf{x}_i | C_j] P[C_j]}{P[\mathbf{x}_i]} = \frac{P[\mathbf{x}_i | C_j] P[C_j]}{\sum_j P[\mathbf{x}_i | C_j] P[C_j]}$$

Maximization Step

■ Input (Know)

$$\begin{aligned} \mathbf{x}_1 &= \{\mathbf{r}_1, \mathbf{g}_1, \mathbf{b}_1\} \\ \mathbf{x}_2 &= \{\mathbf{r}_2, \mathbf{g}_2, \mathbf{b}_2\} \\ &\dots \\ \mathbf{x}_i &= \{\mathbf{r}_i, \mathbf{g}_i, \mathbf{b}_i\} \\ &\dots \end{aligned}$$

■ +

■ Input (Estimation)

■ Classification Results

$$\begin{aligned} &\mathbf{P}[\mathbf{C}_1 \mid \mathbf{x}_1] \\ &\dots \\ &\mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i] \\ &\dots \end{aligned}$$

■ $\Rightarrow$

■ Output

■ Cluster Parameters

$$\begin{aligned} &(\mu_1, \Sigma_1), \mathbf{P}[\mathbf{C}_1] \text{ for } \mathbf{C}_1 \\ &\dots \\ &(\mu_k, \Sigma_k), \mathbf{P}[\mathbf{C}_k] \text{ for } \mathbf{C}_k \\ &\dots \end{aligned}$$

$$\mu_j = \frac{\sum_i \mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i] \mathbf{x}_i}{\sum_i \mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i]}$$

$$\Sigma_j = \frac{\sum_i \mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i] \cdot (\mathbf{x}_i - \mu_j) \cdot (\mathbf{x}_i - \mu_j)^T}{\sum_i \mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i]}$$

$$\mathbf{P}[\mathbf{C}_j] = \frac{\sum_i \mathbf{P}[\mathbf{C}_j \mid \mathbf{x}_i]}{N}$$

Naive Bayesian Classifier

$$P[x | y = c, \theta] = \prod_{j=1}^D P[x_j | y = c, \theta_{jc}]$$

datum instance $x = (a_1, a_2, \dots, a_T)$

$$h_{\text{Naive Bayes}} = \arg \max_h P[h] P[x | h] = \arg \max_h P[h] \prod_t P[a_t | h]$$

■ For each target value (hypothesis) h

$\hat{P}[h] := \text{estimate } P[h]$

■ For each attribute value a_t of each datum instance

$\hat{P}[a_t | h] := \text{estimate } P[a_t | h]$

■ Sample

Outlook $\rightarrow$ Sunny, Overcast, Rain

Temperature $\rightarrow$ Hot, Mild, Cool

Humidity $\rightarrow$ High, Normal, High

Wind $\rightarrow$ Weak, Strong

PlayTennis $\rightarrow$ yes, no

$P[\text{PlayTennis} = \text{yes}] = 9 / 14 = 0.64$

$P[\text{PlayTennis} = \text{No}] = 5 / 14 = 0.36$

$P[\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{yes}] = 3 / 9 = 0.33$

$P[\text{Wind} = \text{Strong} | \text{PlayTennis} = \text{no}] = 3 / 5 = 0.60$

etc.

$P[\text{yes}] P[\text{Sunny} | \text{yes}] P[\text{Cool} | \text{yes}] P[\text{High} | \text{yes}] P[\text{Strong} | \text{yes}] = 0.0053$

$P[\text{no}] P[\text{Sunny} | \text{no}] P[\text{Cool} | \text{no}] P[\text{High} | \text{no}] P[\text{Strong} | \text{no}] = 0.0206$

$\Rightarrow \text{answer} : \text{PlayTennis}[x] = \text{no}$