Assignment-I CS331

Name: Shantanu Wagh
Roll no: 1903217

1) Priors are equal and $f_i(x) = \frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}$ $i=1,2$

$\rightarrow$

a) For class conditional densities to be pdf,

$$\int_{-\infty}^{\infty} f_i(x)\,dx = 1.$$

$$= \int_{-\infty}^{\infty}\left[\frac{1}{\pi b}\cdot\frac{1}{1+\left(\frac{x-a_i}{b}\right)^2}\right]dx.$$

$\frac{x-a_i}{b} \rightarrow t$

$$= \int_{-\infty}^{\infty} \frac{1}{\pi \cancel{b}}\cdot\left(\frac{1}{1+t^2}\right)\cancel{b}\,dt.$$

$$= \frac{1}{\pi}\left[\tan^{-1}(x)\right]_{-\infty}^{\infty} = \frac{1}{\pi}\left[\frac{\pi}{2}+\frac{\pi}{2}\right] = \frac{1}{\pi}\cdot\pi = 1$$

b) We will get boundary when $q_1(x) = q_2(x)$

$$q_1(x) = q_2(x).$$

given, $\{P_1(x) = P_2(x)\}$

$$\cancel{P_1(x)}\, f_1(x) = \cancel{P_2(x)}\, f_2(x)$$

$$\cancel{\frac{1}{\pi b}}\cdot\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2} = \cancel{\frac{1}{\pi b}}\cdot\frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}.$$

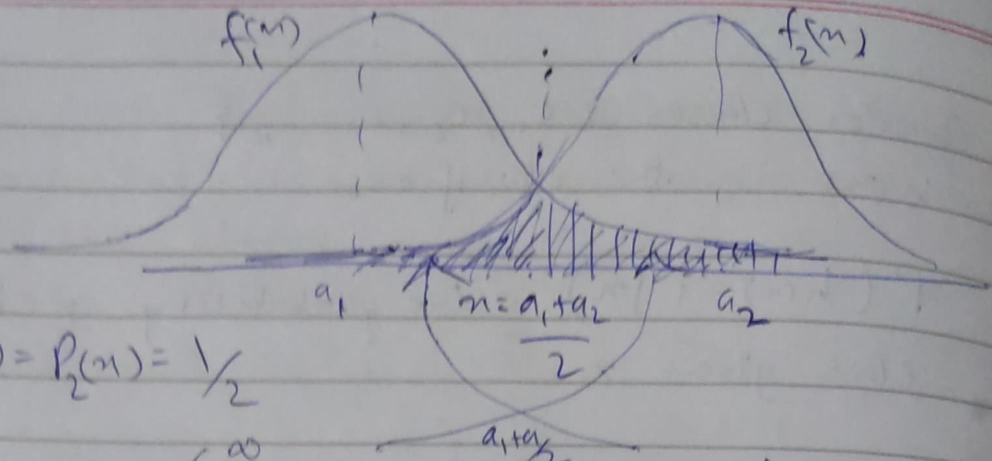
$$(x-a_1)^2-(x-a_2)^2 = 0$$

$$(\cancel{x}-a_1-\cancel{x}+a_2)(x-a_1+x-a_2) = 0$$

$$(a_2-a_1)(2x-a_1-a_2) = 0$$

$$\therefore\ x = \frac{a_1+a_2}{2}.$$ $\leftarrow$ Bayes Decision boundary.

c)



$$P_1(x) = P_2(x) = 1/2$$

$$P_{error} = \frac{1}{2}\left(\int_{\frac{a_1+a_2}{2}}^{\infty} f_1(x)\,dx + \int_{-\infty}^{\frac{a_1+a_2}{2}} f_2(x)\,dx\right)$$

$$= \frac{1}{2\pi b}\left(\int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{1+\frac{(x-a_1)^2}{b^2}}\,dx + \int_{-\infty}^{\frac{a_1+a_2}{2}} \frac{1}{1+\frac{(x-a_2)^2}{b^2}}\,dx\right)$$

Let $\frac{x-a_1}{b} \to t$ & $\frac{x-a_2}{b} \to z$

$$= \frac{1}{2\pi \cancel{b}}\cancel{b}\left(\int_{\frac{a_1-a_2}{2b}}^{\infty} \frac{1}{1+t^2}\,dt + \int_{-\infty}^{\frac{a_1-a_2}{2b}} \frac{1}{1+z^2}\,dz\right)$$

$$= \frac{1}{2\pi}\left(\left[\tan^{-1} t\right]_{\frac{a_1-a_2}{2b}}^{\infty} + \left[\tan^{-1} z\right]_{-\infty}^{\frac{a_1-a_2}{2b}}\right).$$

$$= \frac{1}{2\pi}\left(\frac{\pi}{2} - \tan^{-1}\left(\frac{a_1-a_2}{2b}\right) - \tan^{-1}\left(\frac{a_1-a_2}{2b}\right) + \frac{\pi}{2}\right)$$

$$\boxed{P_{error} = \frac{1}{2\pi}\left(\pi - 2\tan^{-1}\left(\frac{a_1-a_2}{2b}\right)\right).}$$

2) Consider classes $\{c_1, c_2, \ldots c_k\}$
$h(x)$ is the classifiers.

$P(h(x) = i \mid x)$ is the probability of choosing i^{th} class given x.

a) $Risk[(h(x) = i) \mid x]$

→ Risk of choosing i^{th} class when actual class is Y

$\Rightarrow E[L(h(x), Y) \mid X]$

for all possible classes Y

$$\Rightarrow \sum_{j=1}^{k} a_j(x) \, L(h(x), c_j)$$

Posterior probability of getting j^{th} class

Now,

$$L(h(x), c_j) = \begin{cases} 0 & \text{if } h(x) = c_j \\ 1 & \text{if } h(x) \neq c_j \end{cases}$$

$$= \sum_{c_j = h(x)} 0 \cdot a_j(x) + \sum_{c_j \neq h(x)} 1 \cdot a_j(x)$$

Since $\sum a_j(x) = 1$.

Therefore

$Risk[h(x) = i \mid x] = 1 - a_j(x)$ where, $h(x) = c_j$

b) To improve the classifier, we can do the following-
Improve the loss functions as different classes should have different values of loss function.

3) $$L(h(x)=i, y=j) = \begin{cases} 0 & i=j \\ d_r & i=k+1 \\ d_m & \text{otherwise} \end{cases}$$

$d_r \rightarrow$ loss incurred for choosing the $(k+1)^{th}$ class, rejection
& $d_m \rightarrow$ loss incurred for misclassification

We know,
$$Risk[h(x)=i \mid x) = \begin{cases} E[L(h(x)), y) \mid x] \\ d_r \end{cases}$$

$$= \begin{cases} \sum_{i=1} q_i(x) \cdot d_m \\ d_r \end{cases}$$

$$= \begin{cases} d_m(1-q_i(x)) & \text{if misclassification} \\ d_r & \text{if rejection} \end{cases}$$

Minimizing risk for j
$\forall j \quad Risk(h(x)=j \mid x) \geq Risk(h(x)=i \mid x)$
For case 1:
$$\cancel{d_m}(1-q_j(x)) \geq \cancel{d_m}(1-q_i(x))$$

$$\boxed{q_j(x) \leq q_i(x)}$$

For case 2:
$$d_r \geq d_m(1-q_i(x))$$
$$q_i(x) \geq 1-\frac{d_r}{d_m}$$

$$\text{if } \lambda_r = 0 \Rightarrow q_i(x) = 1$$

$$\text{if } \lambda_r > \lambda_m \text{ i.e. } \frac{\lambda_r}{\lambda_m} - 1 > 0$$

$$\Rightarrow 1 - \frac{\lambda_r}{\lambda_m} < 0$$

$$\therefore q_i(x) \geq 0$$

4) a) Exponential distribution.

→ Let $f(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$D = \{x_1, x_2, \ldots x_n\}$ ← data using which we will estimate the value of λ.

likelihood function $L(\theta, D) = \prod_{i=1}^{n} f(x_i|\theta)$

$$L(\lambda, D) = \lambda^n e^{-\lambda(x_1 + x_2 + \ldots x_n)}$$

$$l(\lambda, D) = \sum_{i=1}^{n} (\ln(\lambda) - \lambda x_i)$$

Differentiating and equating to zero.

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\boxed{\lambda = \frac{n}{\sum_{i=1}^{n} x_i}}$$ ← MLE for exponential distribution.

b) Multivariate Gaussian distribution

→ Let X be d dimensional vector of features, μ = mean of d features & Σ = covariance matrix

$$f(x|\theta) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

$\theta = (\mu, \Sigma)$ are the parameters.

Taking log likelihood of the function.

$$l(\mu, \Sigma | D) = \log \prod_{i=1}^{n} f(x^{(i)} | \mu, \Sigma)$$

$$= \log \left(\prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}((x^{(i)}-\mu)^T \Sigma^{-1}(x^{(i)}-\mu))} \right)$$

$$= \sum_{i=1}^{n} \left(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2} \cdot (x^{(i)}-\mu)^T \Sigma^{-1}(x^{(i)}-\mu) \right)$$

taking derivative w.r.t $\theta = (\mu, \Sigma)$

$$\frac{dl}{d\theta} = \begin{bmatrix} dl/d\mu \\ dl/d\Sigma \end{bmatrix}$$ and equating to 0.

$$\frac{dl}{d\mu} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \Sigma^{-1}(\mu - x^{(i)}) = 0$$

$$\boxed{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}}$$

$$\frac{dl}{d\Sigma} = 0 \quad \Rightarrow \quad n\Sigma - \sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

$$\boxed{\hat{\Sigma} = \frac{1}{n}\left(\sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T\right)}$$

5) $f(D|\mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

the prior for above likelihood is given by,

$$f(\mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) \cdot e^{-\frac{1}{2\sigma}(\mu-\mu_0)}$$

$\therefore$ posterior will be

$$f(\mu|D) = \frac{f(D|\mu) \cdot f(\mu)}{\int f(D|\mu) \cdot f(\mu)\, d\mu}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right)$$

$$= k \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right).$$

Taking log likelihood.

$$l(\mu|D) = \ln k - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2$$

Differentiating w.r.t 'μ' & then equating to 0.

$$\frac{-1}{2\sigma^2}\left(2n\mu - 2\sum_{i=1}^{n}x_i\right) = \frac{(\mu-\mu_0)}{\sigma_0^2}$$

$$\boxed{\mu = \frac{\left(\sigma_0^2 \sum_{i=1}^{n} x_i + \sigma^2\mu_0\right)}{(\sigma_0^2+\sigma^2)}}$$

$$k\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 - \frac{n}{2\sigma^2}\mu^2 + \frac{1}{2\sigma^2}2\mu\sum_{i=1}^{n}x_i - \frac{1}{2\sigma_0^2}\mu^2 - \frac{1}{2\sigma_0^2}\mu_0^2 + \frac{2\mu\mu_0}{2\sigma_0^2}\right)$$

$$= k\exp(a\mu^2 + b\mu + c)$$

$$\int k\exp(a\mu^2+b\mu+c) = 1$$

$$a = \frac{-n}{2\sigma^2} - \frac{1}{2\sigma_0^2}$$

$$b = \frac{1}{2\sigma^2} \times 2\sum_{i=1}^{n}x_i + \frac{2\mu_0}{2\sigma_0^2}$$

$$c = \frac{-1}{2\sigma^2}\sum x_i^2 - \frac{1}{2\sigma_0^2}\mu_0^2$$

Date
Page

As posterior & prior have the same form of likelihood, the distribution will be normal and mean and mode will be same.

so,

$$\boxed{\mu = \frac{\sigma_0^2 \sum_{i=1}^{n} x_i + \sigma^2 \mu_0}{(\sigma_0^2 + \sigma^2)}}$$

6) a)

$\rightarrow$ $f(x|\theta) = \alpha_1 f(x|\theta_1) + \alpha_2 f(x|\theta_2) + \alpha_3 f(x|\theta_3) + \dots$
$\dots \alpha_t f(x|\theta_t)$

Let the hidden variable be $z_i = \{z_{i1}, z_{i2}, z_{i3} \dots z_{it}\}$

$$P(z_{ij} = 1) = \alpha_j \ \forall i$$

$$f(x_i | z_{ij} = 1) = \phi(x_i | \theta_j) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right)$$

$$f(x_i | z_i, \theta) = \prod_{j=1}^{t} (\alpha_j \phi(x_i | \theta_j))^{z_{ij}}$$

The complete likelihood,

$$f(x, z | \theta) = \sum_{i=1}^{n} \left(\prod_{j=1}^{t} (\alpha_j \phi(x_i | \theta_j))^{z_{ij}} \right)$$

Taking log likelihood,

$$\ln(f(x, z|\theta)) = \sum_{i=1}^{n} \left(\sum_{j=1}^{t} z_{ij} \ln(\alpha_j \phi(x_i | \theta_j)) \right)$$

E step,

Expected value of $\ln(f(x, z|\theta))$ over Z for any specific θ'.

$$E[z_{ij} | x, \theta'] = P[z_{ij} = 1 | x, \theta'] = P[z_{ij} = 1 | x_j, \theta']$$

$$Q(\theta, \theta^{(k)}) = \sum_{i=1}^{n} \left[\sum_{i=1}^{t} E[z_{ij} | x, \theta^{(k)}] \ln(\lambda_j \phi(x_i | \theta_j)) \right]$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{t} \gamma_{ij}(\theta^{(k)}) \right) \ln\left(\lambda_j \phi(x_i | \theta_j)\right)$$

$$\gamma_{ij}(\theta^{(k)}) = \frac{\lambda_j \phi(x_i | \theta_j^{(k)})}{\sum_{j=1}^{t} \lambda_j \phi(x_i | \theta_j^{(k)})}$$

$$= \frac{\lambda_j \frac{1}{(\sigma_j)^k \sqrt{2\pi}} \exp\left(\frac{-(x_i - \mu_j^{(k)})^2}{2(\sigma_j^{(k)})^2} \right)}{\sum_{j=1}^{t} \lambda_j \frac{1}{(\sigma_j)^k \sqrt{2\pi}} \exp\left(\frac{-(x_i - \mu_j^{(k)})^2}{2(\sigma_j^{(k)})^2} \right)}$$

classmate
Date
Page

Now M step,
We find $\theta^{(k+1)}$ that maximizes over θ.

$$Q(\theta, \theta^{k}) = \sum_{i=1}^{n} \sum_{j=1}^{t} r_{ij}(\theta^{(k)}) \left[\ln \alpha_j - \ln(\sigma_j \sqrt{2\pi}) - \frac{(x_i - \mu_j)^2}{2\sigma_j^2} \right]$$

Differentiating w.r.t. μ_j and equating to 0.

$$\sum_{i=1}^{n} r_{ij}(\theta^{(k)}) \frac{(x_i - \mu_j)}{\sigma_j^2} = 0$$

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^{n} r_{ij}(\theta^{(k)}) x_i}{\sum r_{ij}(\theta^{(k)})}$$

Differentiating w.r.t. σ_j & equating to 0

$$\sum_{i=1}^{n} r_{ij}(\theta^{(k)}) \left[\frac{-1}{\sigma_j} + \frac{(x_i - \mu_j)^2}{\sigma_j^3} \right] = 0.$$

$$(\sigma_j^2)^{(k+1)} = \frac{\sum_{i=1}^{n} r_{ij}(\theta^{(k)})(x_i - \mu_j^{(k)})^2}{\sum_{i=1}^{n} r_{ij}(\theta^{(k)})}$$

Differentiating Lagrange w.r.t α_j & equating to 0.

$$l(f(x|\theta, z)) + \eta\left(\sum_{j=1}^{t} \alpha_j - 1\right).$$

where η is the lagrange Multiple.

$$\boxed{\sum_{i=1}^{n} \frac{r_{ij}}{\alpha_j} + \eta = 0} \quad -①$$

Similarly for all α_j's we will get similar eq^n.

$$\eta = n \sum_{i=1}^{t} \alpha_j$$

Combining eqn for all λ_j we will get

$$\eta \sum_{j=1}^{t} \lambda_j = -\sum_{i=1}^{n} \sum_{j=1}^{t} (r_{ij})$$

$$\therefore \eta = -n \qquad \left\{ \sum_{i=1}^{n} \sum_{j=1}^{t} (r_{ij}) = n \right\}$$

$$\therefore \lambda_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} r_{ij}^{(k)}$$

$$\& \quad r_{ij}^{(k+1)} = \frac{\lambda_j^{(k+1)} \phi(x_i | \theta_j^{(k)})}{\sum_{j=1}^{t} \lambda_j \phi(x_i | \theta_j^{(k)})}$$

b) Mixture of bernoulli

$\rightarrow$ $f(x|\theta) = \lambda_1 f(x|\theta_1) + \lambda_2 f(x|\theta_2) + \cdots \lambda_t f(x|\theta_t)$

Let the hidden variable be,

$$z_i = \{z_{i1}, z_{i2}, \cdots z_{it}\}$$

$$P(z_{ij} = 1) = \lambda_i \quad \forall i$$

$$f(x_i = z_{ij}) = \phi(x_i | \theta_j) = p_j^{x_i} (1 - p_j)^{1 - x_i}$$

$$f(x_i | z_i, \theta) = \prod_{j=1}^{t} (\lambda_i \phi(x_i | \theta_j))^{z_{ij}} \quad \left[\begin{array}{l} \text{complete} \\ \text{density function} \end{array} \right]$$

Complete likelihood,

$$f(x, z | \theta) = \prod_{i=1}^{n} \left(\prod_{j=1}^{t} (\lambda_j \phi(x_i | \theta_j))^{z_{ij}} \right)$$

Taking log,

$$\ln(f(x, z | \theta)) = \sum_{i=1}^{n} \left(\sum_{j=1}^{t} z_{ij} \ln(\lambda_j \phi(x_i | \theta_j)) \right)$$

E step,

$$Q(\theta, \theta^{(k)}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{t} E[z_{ij} | x, \theta^{(k)}] \ln(\alpha_j \phi(x_i | \theta_j)) \right).$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{t} r_{ij}(\theta^{(k)}) \ln \alpha_j (\phi(x_i | \theta_j)) \right)$$

where,

$$r_{ij}(\theta^{(k)}) = \frac{\alpha_j^{(k)} \phi(x_i | \theta_j^{(k)})}{\sum_{j=1}^{t} \alpha_j^{(k)} \phi(x_i | \theta_j^{(k)})} = \frac{\alpha_j (p_j^{(k)})^{x_i} (1 - p_j^{(k)})^{1-x_i}}{\sum_{j=1}^{t} \alpha_j^{(k)} (p_j^{k})^{x_i} (1 - p_j^{k})^{1-x_i}}$$

M step,

$$Q(\theta, \theta^{k}) = \sum_{i=1}^{n} \sum_{j=1}^{t} r_{ij}(\theta^{(k)}) [\ln \alpha_j + x_i \ln p_j + (1 - x_i) \ln(1 - p_j)]$$

Differentiating w.r.t p_j & equating to 0.

$$\sum_{i=1}^{n} r_{ij}(\theta^{k}) \left[\frac{x_i}{p_j} - \frac{1 - x_i}{(1 - p_j)} \right] = 0$$

$$\therefore p_j^{(k+1)} = \frac{\sum_{i=1}^{n} r_{ij}(\theta^{(k)}) x_i}{\sum_{i=1}^{n} r_{ij}(\theta^{k})}$$

Differentiating Lagrange w.r.t α_j & equating to 0.

$$l(f(x | \theta, z)) + \eta \left(\sum_{i=1}^{t} \alpha_j - 1 \right)$$ where η is Lagrange Multiplier.

$$\therefore \sum_{i=1}^{n} \frac{r_{ij}}{\alpha_j} + n = 0$$

Similar to previous part $\eta = -n$

$$\boxed{\therefore \alpha_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} r_{ij}^{(k)}}$$

&

$$r_{ij}^{(k+1)} = \frac{\alpha_j^{(k+1)} \phi(x_i | \theta_j^{(k+1)})}{\sum_{j=1}^{t} \alpha_j^{(k+1)} (\phi(x_i | \theta_j^{(k+1)}))}.$$

classmate

Date ______

Page ______

7) Derive the MAP estimate of a Bernoulli distribution
→ based on n iid samples. (Conjugate prior: Beta distribution)

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \qquad p \in [0,1],\ a,b \geq 1$$

Beta(p|a,b)

By differentiating we can easily show that its mode is at $\frac{a-1}{a+b-2}$

$$f(p|D) = k f(D|p) f(p)$$

$$= k_1 p^{\sum x_i}(1-p)^{n-\sum x_i} p^{a-1}(1-p)^{b-1}$$

$$= k_1 p^{\sum x_i + a - 1}(1-p)^{n+b-\sum x_i - 1}$$

Hence the posterior is Beta$(\sum x_i + a, n+b-\sum x_i)$

Now we want the MAP estimate,

we know, mode of Beta(a,b) is $\frac{a-1}{a+b-2}$

Hence MAP estimate (mode of posterior density) is given by

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i + a - 1}{n+a+b-2}$$