Assignment 2

Sunday December 25, 2022 12:17 PM

Q.1 Gaussian distribution → Gaussian distribution

→ To calculate the mean of random variable and we data for that particular class

given information:- total datapoints - 20 points

class 1 → 6 datapoints

class 2 → 4 datapoints

π_1 → mean for class 1 for feature X

$\pi_1 = \frac{0.0+0.0+0.0+0.0+0.0+0.0}{6} = 0$

$\pi_2 = \frac{1+1+1+1+1+1}{6} = \frac{6}{6} = 1.0$

$\Rightarrow \boxed{\pi_1 = \begin{bmatrix} \pi_{11} \\ \pi_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 1.0 \end{bmatrix}}$ (1)

similarly we calculate π_2

$\pi_{21} = \frac{0.5+0.5+0.5+0.5}{4} = \frac{2.0}{4} = 0.5$

$\pi_{22} = \frac{1+1+1+1+1}{4} = \frac{4}{4} = 1.0$

$\Rightarrow \boxed{\pi_2 = \begin{bmatrix} \pi_{21} \\ \pi_{22} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}}$ (2)

Now, we calculate the covariance of the two classes

$\boxed{\text{Covariance} = \frac{(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}}$ where x_i → observation of one feature, y_i → observation of another feature, $\bar{x}, \bar{y}$ → respective means } sample covariance

i.e class

∴ for class 1

$Cov(X,Y)_1 = \frac{(0.0-1.0)\cdot(1.0-1.375) + (0.0-1.0)(1.0-1.375) + (1.0-1)(1.0-1.375) + (1.0-1)(1.0-1.375)}{3}$

$Cov(X,X)_1 = \frac{(0.0\times 0.0125) + (0.0\times 0.0125) + 0\times 0.0125 + 0\times 0.0125}{3}$

$= \frac{0.0125 + 0.0125 + 0.0125 + 0.0125}{3}$

$\boxed{Cov(X,Y)_1 = \frac{0.05}{3} = 0.01666}$ (3)

$Cov(Y,X)_1 = Cov(X,Y)_1 = \frac{(0.0-1.0)\cdot(0.0-1.0) + (0-1.0)^2 + (0-1.0)^2 + (0-1.0)^2}{3}$

$= \frac{0.0 + 0.0 + 0.0 + 0.0}{3}$

$\boxed{Var(X,X)_1 = \frac{0.3}{3} = 0.0333}$ (4)

$Cov(Y,Y)_1 = \frac{0.125^2 + 0.0125^2 + 0.0125^2 + 0.125^2}{3} = \frac{0.0156125 + 0.000125 + 0.000125 + 0.0156125}{3}$

$\boxed{Var(Y,Y)_1 = \frac{0.03125}{3} = 0.01083} - (5)$

∴ covariance - covariance matrix for class 1 is $\boxed{Cov(X,Y)_1 = \begin{bmatrix} 0.0333 & 0.01666 \\ 0.0166 & 0.01083 \end{bmatrix}}$ (6)

Similarly, we can calculate variance and covariance for class 2, given we know that $\pi_2 = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}$

$Cov(X,Y)_2 = \frac{(0.5-0.5)(1.1-1.025) + (0.5-0.5)(1.1-1.025) + (0.45-0.5)(1.1-1.025) + (0.75-0.5)(0.9-1.025)}{3}$

$= \frac{(0.1\cdot 0.025) + (0.1\cdot 0.025) + (-0.05)\cdot(-0.025) + 0.25\cdot(-0.125)}{3}$

$= \frac{0.0025 + 0.0025 - 0.00125 - 0.0625}{3} = \frac{0}{3}$

$\boxed{Cov(X,Y)_2 = 0}$ (7)

$Cov(X,X)_2 = \frac{0.1^2 + 0.1^2 + 0.05^2 + 0.05^2}{3} = \frac{0.01 + 0.01 + 0.0025 + 0.0025}{3}$

$Cov(X,X)_2 = \frac{0.025}{3} = 0.008333$

$\therefore \boxed{Var(X,X)_2 = 0.008333}$ (8)

$Cov(Y,Y)_2 = \frac{0.075^2 + 0.025^2 + 0.075^2 + 0.125^2}{3}$

$= \frac{0.005625 + 0.000625 + 0.005625 + 0.015625}{3}$

$= \frac{0.0275}{3} = 0.009166$

$\therefore \boxed{Var(Y,Y)_2 = 0.009166}$ (9)

$\therefore \boxed{Cov(X,Y)_2 = \begin{bmatrix} 0.0833 & 0 \\ 0 & 0.00916 \end{bmatrix}}$ (10)

Q.3 Use MLE to derive the equation for mean & variance assuming pdf (or likelihood) is modelled by 1-D gaussian.

Ans:- given information - pdf is modelled by 1-D gaussian.

The formula for n-dimension gaussian pdf is

$$\boxed{p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n \{\Sigma\}^{1/2}}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right]}$$

When we only have 1 dimension, the above equation turns to

$\Sigma \Rightarrow \sigma^2$; $\vec{\mu} \Rightarrow \mu$, $\vec{x} \Rightarrow x$, $\Sigma^{-1} \Rightarrow \frac{1}{\sigma^2}$

Using the above notations, the new pdf is

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]}$$

objective of MLE :- to find the set of parameters so that it maximizes the likelihood of observing the data

Here the set of parameters $\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$

$pdf\ p(x;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$

We read the above equation as probabability of obtaining x given unknown parameters θ

In the likelihood function, x is fixed and θ is a function of x

$\mathcal{L}(\theta|x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$

Suppose we have $\{x_1, x_2, x_3, x_4 \dots x_n\}$ training samples

We make an important assumption

All the samples are drawn independently

$\therefore p(D;\vec{\theta}) = p(x_1, x_2, x_3 \dots x_n;\vec{\theta}) = p(x_1 \cap x_2 \cap x_3 \cap x_4 \dots \cap x_n;\vec{\theta})$

$= p(x_1;\vec{\theta}) \cdot p(x_2;\vec{\theta}) \cdot p(x_3;\vec{\theta}) \cdots p(x_n;\vec{\theta})$

$p(D;\vec{\theta}) = \prod_{i=1}^{n} p(x_i;\vec{\theta})$

Applying log to both sides

$\mathcal{L}(\vec{\theta}) = \ln p(D;\vec{\theta}) = \sum_{i=1}^{n} \ln p(x_i;\vec{\theta})$

We have to find $\hat{\theta}$ that maximizes the above function

$\hat{\theta} = \underset{\vec{\theta}}{\arg\max}\ \mathcal{L}(\vec{\theta}) = \underset{\vec{\theta}}{\arg\max} \sum_{i=1}^{n} \ln p(x_i;\vec{\theta})$

Now we know that

$p(x;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$

$\mathcal{L}(\vec{\theta}) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]\right)$

$\mathcal{L}(\vec{\theta}) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln 2\pi - \ln\sigma - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$

$\mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$

$\mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2 - 2x_i\mu + \mu^2}{\sigma^2}$

$\mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}2x_i\mu + \sum_{i=1}^{n}\mu^2\right\}$

$\mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}2x_i\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mu^2$

To find the best parameters that maximize the likelihood of obtaining the data, we differentiate the above equation w.r.t. the parameters and equate it to 0

Here parameters $\vec{\theta} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$

i) Equating the partial derivative with respect to (w.r.t) μ equal to 0

$\frac{\partial \mathcal{L}(\vec{\theta})}{\partial \mu} = \frac{\partial}{\partial \mu}\left\{-\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}2x_i\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mu^2\right\}$

$= \frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^{n}2x_i - \frac{\cancel{2}}{\cancel{2}\sigma^2}\sum_{i=1}^{n}\mu = \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu = 0$

$n\mu = \sum_{i=1}^{n}x_i \Rightarrow \boxed{\mu = \frac{\sum_{i=1}^{n}x_i}{n}}$

Best estimate of mean according to MLE is mean of samples

ii) Equating the partial derivative by σ, equal to 0

$\frac{\partial \mathcal{L}(\vec{\theta})}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left[-\frac{1}{2}\sum_{i=1}^{n}\ln 2\pi - \sum_{i=1}^{n}\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}2x_i\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mu^2\right]$

$= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}x_i^2 - \frac{1}{\sigma^3}\sum_{i=1}^{n}2x_i\mu + \frac{1}{\sigma^3}\sum_{i=1}^{n}\mu^2 = 0$

$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$

$\frac{1}{\sigma^{\cancel{3}\,2}}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{n}{\cancel{\sigma}}$

$\therefore \boxed{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$

Therefore the best estimate of σ^2 is the variance of the sample according to MLE.