# ASSIGNMENT-1

Name: Amrita Sundari. V

1. A secret government agency has developed a scanner which determines whether a person is a terrorist. The scanner is fairly reliable; 95% of all scanned terrorists are identified as terrorists, and ~~95%~~ an informant tells the agency that exactly one passenger of 100 aboard an aeroplane in which you are seated is a terrorist. The Police haul off the plane the first person for which the scanner tests positive. What is the probability that this person is a terrorist? Additionally, if the police were to scan all passengers, how many positive detections should we expect?.

Let T - be the event when the scanner identifies a person as a terrorist

T' - be the event when the scanner identifies a person as not a terrorist

A - Set of terrorists

A' - Set of Citizens

Given:

* $P(T/A) = 0.95$ $\qquad P(T'/A') = 0.95$

* $P(A) = 1/100 = 0.01$ $\quad \Rightarrow P(A') = 0.99$

To find:

Probability the scanned person is a terrorist when the scanner identifies him ~~not~~ as a terrorist.

i.e $P(A/T)$.

By Baye's rule

$$P(A/T) = \frac{P(T/A).P(A)}{P(T/A).P(A) + P(T/A').P(A')}$$

$$= \frac{0.95 \times 0.01}{(0.95 \times 0.01) + (0.05 \times 0.99)}$$

$$\left[ \because P(T/A') = 1 - P(T'/A') \right]$$
$$= 1 - 0.95$$
$$= 0.05.$$

$$= 0.161$$

(b) what is the expected value of the positive detection of the scanner.

i.e We can model the no. of positive detections as a Binomial Random variable. X - no. of +ve detection.

$$E[x] = \sum_{x=1}^{100} x \cdot {}^{n}C_{x} \cdot p^{x} \cdot (1-p)^{n-x}$$

$$E[x] = np.$$

$$= 100 \times p[T]$$

(where $p[T]$ - Total probability that the scanner shows positive)

$$\because p[T] = P[T/A] \cdot p(A) + P(T/A') \cdot P(A')$$
$$= (0.95 \times 0.01) + (0.05 \times 0.99)$$
$$= 0.059$$

$$\Rightarrow E[x] = 100 \times 0.059$$
$$= 5.9$$

$\therefore$ No. of positive detection $\simeq$ 6.

2. The Poisson distribution is a discrete distribution on the non-negative integers with

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0, 1, 2 \ldots$$

You are given a sample of $n$ observations $x_1, \ldots x_n$ independently drawn from this distribution. Determine the Maximum Likelihood Estimator of the poisson parameter $\lambda$.

Likelihood function is given by

$$L(\lambda \mid x_1, x_2 .. x_n) = \prod_{j=1}^{n} \frac{e^{-\lambda}\lambda^{x_j}}{x_j!} \quad \text{—— (1)}$$

we have to $\lambda$ that maximises $L(\lambda/X)$.

Therefore, $\frac{dL}{d\lambda} = 0$ is an equation that $\lambda$ has to satisfy.

$\Rightarrow$ Taking $\ln$ of $L(\lambda/x)$, we get

$$L = \frac{e^{-\lambda n}\lambda^{\sum_{j=1}^{n} x_j}}{\prod_{j=1}^{n} x_j!}$$

$$\ln L = -\lambda n + \left[\sum_{j=1}^{n} x_j\right]\left[\ln \lambda\right] - \left[\sum_{j=1}^{n} \ln x_j!\right]$$

Taking a derivative on both sides and equating it to zero we get.

$$\frac{1}{L} \cdot \frac{dL}{d\lambda} = -n + \frac{1}{\lambda}\left(\sum_{j=1}^{n} x_j\right) = 0$$

$$\Rightarrow \boxed{\lambda = \frac{\sum_{j=1}^{n} x_j}{n}}$$

3. Consider a classifier that makes R correct classifications and W wrong classifications. Is the classifier better than random guessing? Let D represent the fact that there are R right and W wrong answers. Assume also that the classifications are i.i.d.

1. Show that under the hypothesis that the data is generated purely at random, the likelihood is

$$P(D/H_{random}) = 0.5^{R+W}.$$

When we consider the hypothesis to be random, each of the outcome has probability of 0.5 occurance. (Since it can be either correct or wrong.

$$\Rightarrow P(D/H_{random}) = \prod_{k=1}^{R}(0.5) \cdot \prod_{l=1}^{W}(0.5)$$

$$= (0.5)^R \cdot (0.5)^W$$

$$= (0.5)^{R+W}.$$

2. Define $\theta$ to be the probability that the classifier makes an error. Then.

$$P(D/\theta) = \theta^R (1-\theta)^W$$

Now consider

$$P(D/H_{non\text{-}random}) = \int_\theta P(D/\theta) \cdot p(\theta)$$

Show that for a Beta prior $P(\theta) = B(\theta|a,b)$

$$P(D/H_{non\text{-}random}) = \frac{B(R+a, W+b)}{B(a,b)}$$

$$P(D/H_{non\text{-}random}) = \int_\theta P(D/\theta) \cdot P(\theta)$$

$$= \int \theta^R (1-\theta)^W \cdot P(\theta)$$

$$= \int_{\theta} \theta^{R}(1-\theta)^{W}\, B(\theta/a,b)$$

$$\because\ B(\theta/a,b) = \frac{1}{B(a,b)} \cdot \theta^{a-1}(1-\theta)^{b-1}$$

$$\text{and } B(a,b) = \int \theta^{a-1}(1-\theta)^{b-1} \cdot d\theta$$

$$= \int_{\theta} \theta^{R}(1-\theta)^{W} \frac{1}{B(a,b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1} \cdot d\theta$$

$$= \frac{1}{B(a,b)} \underbrace{\int_{\theta} \theta^{R+a-1} \cdot (1-\theta)^{W+b-1} \cdot d\theta}_{B(R+a,\, W+b)}$$

$$= \frac{B(R+a, W+b)}{B(a,b)}$$

3. Considering the random and non-random hypothesis as a priori equally likely, show that

$$P(H_{random}/D) = \frac{0.5^{R+W}}{0.5^{R+W} + \frac{B(R+a, W+b)}{B(a,b)}}$$

By Baye's rule.

$$P(H_{random}/D) = \frac{P(D/H_{random}) \cdot P(H_{random})}{P(D/H_{random}) \cdot P(H_{random}) + P(D/H_{non\text{-}rand}) \cdot P(H}$$

$\because$ given that $P(H_{random}) = P(H_{non\text{-}random})$

$$\Rightarrow = \frac{P(D/H_{random})}{P(D/H_{random}) + P(D/H_{non\text{-}random})}$$

Scanned by CamScanner

standard deviation in the number of errors

$$= \frac{0.5^{R+W}}{(0.5)^{R+W} + \frac{B(R+a,\, W+b)}{B(a,b)}}$$

4. For a flat prior $a = b = 1$, Compute the probability that for 10 correct and 12 incorrect classifications, the data is from a purely random distribution. Repeat this for 100 correct & 120 incorrect classifications.

(a)
$$P(H_{random}/D) = \frac{0.5^{R+W}}{0.5^{R+W} + \frac{B(R+a,\, W+b)}{B(a,b)}}$$

$$= \frac{0.5^{22}}{0.5^{22} + \frac{B(11,13)}{B(1,1)}}$$

$$= \frac{0.5^{22}}{0.5^{22} + \frac{10!\, 12!}{23!}}$$

$$= 0.78$$

(b)
$$P(H_{random}/D) = \frac{0.5^{220}}{(0.5)^{220} + B(101, 121)}$$

$$= \frac{0.5^{220}}{0.5^{220} + \frac{100!\, 120!}{221!}}$$

$$= 0.8275$$

Scanned by CamScanner

5) show that ~~the~~ standard deviation in the number of errors of a random classifier is $0.5\sqrt{R+W}$ and relate this to the above computation.

Variance for a binomial distribution is given by

$$\sigma^2 = np(1-P)$$

where $n$ - no of attempts $= R+W$.

$P$ - probability of error $= 0.5$ (for Random classifier)

$$\Rightarrow \sigma^2 = (R+W)(0.5)(1-0.5)$$

$$= (0.5)^2 (R+W)$$

Standard deviation $\sigma = (0.5)\sqrt{R+W}$.

In the previous part

* $R=10$, $W=12$

Standard deviation $= 0.5\sqrt{22}$

$= 2.345$.

* $R=100$ $W=120$.

Standard deviation $= 0.5\sqrt{220}$.

$= 7.416$

Scanned by CamScanner

4. For a novel input $x$, a predictive model of the class $C$ is given by $p(C=1/x) = 0.7$, $p(C=2/x) = 0.2$, $p(C=3/x) = 0.1$. The Corresponding Utility matrix $U(C^{true}, C^{pred})$ has elements

$$\begin{pmatrix} 5 & 3 & 1 \\ 0 & 4 & -2 \\ -3 & 0 & 10 \end{pmatrix}$$

In terms of maximal expected utility, which is the best decision to take?

$$U_{ij} = U(C^{true} = i, C^{pred} = j)$$

$$U(c(x^*)) = \sum_{c^{true}} U(c^{true}, c(x^*)) \cdot p(c^{true}/x^*)$$

$$U(c(x^*) = 1) = U(c=1, c=1) \cdot p(c=1/x) + U(c=2, c=1) \cdot p(c=2/x) + U(c=3, c=1) \cdot p(c=3/x)$$

$$= 5 \times 0.7 + 0 \times 0.2 + (-3) \times 0.1$$

$$= 3.5 - 0.3$$

$$= 3.2$$

$$U(c(x^*) = 2) = U(C=1, C=2) \cdot p(c=1/x) + U(c=2, c=2) \cdot p(c=2/x) + U(c=3, c=2) \cdot p(c=2/x)$$

$$= 3 \times 0.7 + 4 \times 0.2 + 0 \times 0.1$$

$$= 2.1 + 0.8$$

$$= 2.9$$

$$U(c(x^*) = 3) = U(c=1, c=3) \cdot p(c=1/x) + U(c=2, c=3) \cdot p(c=2/x) + U(c=3, c=3) \cdot p(c$$

$$= 1 \times 0.7 + (-2) \times 0.2 + 10(0.1)$$

$$= 0.7 - 0.4 + 1$$

$$= 1.3$$

In terms of maximal expected utility, the best decision to take is Class 1

Scanned by CamScanner

5. Consider datapoints generated from 2 different classes. Class 1 has the distribution $p(x/c=1) \sim N(x \mid m_1, \sigma^2)$ and class 2 has the distribution $p(x/c=2) \sim N(x/m_2, \sigma^2)$. The Prior probability of each class are $p(c=1) = p(c=2) = 1/2$. Show that posterior probability $p(c=1/x)$ is of the form.

$$p(c=1|x) = \frac{1}{1+\exp(-(ax+b))}$$ and determine a & b

in terms of $m_1, m_2$ and $\sigma^2$

$$p(x/c=1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m_1)^2}{2\sigma^2}}$$

$$p(x/c=2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m_2)^2}{2\sigma^2}}$$

$$p(c=1/x) = \frac{p(x/c=1).p(c=1)}{p(x/c=1).p(c=1) + p(x/c=2).p(c=2)}$$

$$= \frac{e^{-\frac{(x-m_1)^2}{2\sigma^2}}}{e^{-\frac{(x-m_1)^2}{2\sigma^2}} + e^{-\frac{(x-m_2)^2}{2\sigma^2}}}$$

Dividing the numerator & denominator by $e^{-\frac{(x-m_1)^2}{2\sigma^2}}$

$$= \frac{1}{1 + \exp\left[-\frac{(x-m_2)^2}{2\sigma^2} + \frac{(x-m_1)^2}{2\sigma^2}\right]}$$

$$= \frac{1}{1+\exp\left[\frac{m_1^2 - m_2^2 + 2(m_2-m_1)x}{2\sigma^2}\right]}$$

$$= \frac{1}{1+\exp\left[-\frac{[(m_2^2 - m_1^2) + 2(m_1-m_2)x]}{2\sigma^2}\right]}$$

Scanned by CamScanner

This is of the form $\dfrac{1}{1+\exp(-(ax+b))}$

where

$$\boxed{b = \frac{m_2^2 - m_1^2}{2\sigma^2} \quad \text{and} \quad a = \frac{m_1 - m_2}{\sigma^2}}$$

—— × ——