D. Arvind Sai Dooda A20553046

Assignment-2

Problem 1: Independence and law of Total Probability

let x y z all be binary variables, taking either 0 or 1. Assume y and z are independent and $P(Y=1)=0.9$ while $P(z=1)=0.8$ Further, $P(x=1 \mid Y=1, z=1)=0.6$ and $P(x=1 \mid Y=1, Z=0)=0.1$, and $P(X=1 \mid Y=0)=0.2$

i. Compute $P(x=1)$.

using law of total probability.

$P(X=1) = P(X=1 \mid Y=1, Z=1)\, P(Y=1, Z=1) +$
$P(X=1 \mid Y=1, Z=0)\, P(Y=1, Z=0) + P(X=1 \mid Y=0)\, P(Y=0)$

$P(Y=1, Z=1) = P(Y=1)\, P(Z=1)$ ($\because$ Y and Z are independent)

$= 0.9 * 0.8$

$= 0.72$

$P(Y=1, Z=0) = P(Y=1)\, P(Z=0)$

$= 0.9 * 0.2$

$= 0.18$

$P(Y=0) = 1 - P(Y=1)$

$= 1 - 0.9$

$= 0.1$

now, we can check

$P(X=1) = 0.6 * 0.72 + 0.1 * 0.18 + 0.2 * 0.1$

$= 0.432 + 0.018 + 0.02$

$= 0.47$

2. compute the expected value $E[Y]$

$\Rightarrow$

the formulae:

$E[Y] = \Sigma y_i P(y_i)$ or $y P(Y=y)$

$\because$ The value of y are 0 & 1.

$P(0) \Rightarrow P(y=0) = 1-P(y=1)$

$P(y=1) = 0.9$.

$P(y=0) = 1-P(y=1)$

$= 1-0.9$

$= 0.1$

$E[Y] = (0)\,P(y=0) + (1)\,P(y=1)$

$= 0(0.1) + 1(0.9)$

$= 0+0.9$

$\boxed{E[Y] = 0.9}$

3. SPPose that instead of Y attaining value 0 and 1, it takes one of Two values 115 and 20. where P(Y = 115) = 0.9. compute value E[Y].

$\therefore E[Y] = (115 * 0.9) + (20 * (1 - 0.9))$

$E[Y] = 103.5 + 2$

$E[Y] = 105.5$

So, when we even try with different value for Y, at last the the value expected for E[Y] remains 105.5

Problem 2 Bayes Rule

Alex owns a retail store for selling phones. Phones manufactured at three different factories A, B, C where factory A, B and C produces 20%, 30% and 50% of the phone being sold at Alex's store. The probability of defective stores. A, B and C are 2%, 1% and 0.05%.

1. what is the Probability of a Phone being defective?

P(Defective) = P(Defective | Factory A) * P(Factory A) +
P(Defective | Factory B) * P(Factory B) +
P(Defective | Factory C) * P(Factory C)

P(Defective | Factory A) = 2% = 0.02

P(Defective | Factory B) = 1% = 0.01

P(Defective | Factory C) = 0.05% = 0.0005

P(Factory A) = 20% = 0.20

P(Factory B) = 30% = 0.30

P(Factory C) = 50% = 0.50

Now, we can add in the formula.

P(Defective) = (0.02 * 0.20) + (0.01 * 0.30) +
(0.0005 * 0.50)
= 0.004 + 0.003 + 0.00025

P(Defective) = 0.725%

2. what is the Probability of that this defective Phone is manufactured at factory A?

∴ Probability of defective Phone manufactured at factory A.

$P(\text{Factory } A \mid \text{Defective}) = (P(\text{Defective} \mid \text{Factory } A) \cdot P(\text{Factory } A)) / P(\text{Defective})$

$P(\text{Factory } A \mid \text{Defective}) = (0.02 * 0.20) / 0.00725$

$= 0.004 / 0.00725$

$= 0.5497$ or 54.97

$\therefore 0.5497 \times 1000$

$= 54.97\%$

The Probability that the defective phone is manufacture by factory A is 54.97%.

3. what is the Probability that this defective Phone is manufactured by factory B.?

now, the Probability of defective phone is from factory B; we are calculating..

P(Factory B | Defective) = (P(Defective | Factory B) * P(Factory B)) | P(Defective)

P(Factory B | Defective) = (0.01 * 0.30) / 0.00725

∴ The above values done at Problem 1

= 0.003 / 0.00725

= 0.4138 or 41.38%.

∴ 0.4138 × 100

= 41.38.

The Probability that the defective phone is manufacture by factory B is : 41.38%.

4. What is the Probability that this defective Phone is manufactured at factory C?

The Probability for the defective phone is manufactured at factory C.

$P(\text{Factory } C \mid \text{Defective}) = (0.0005 * 0.50) / 0.00725$

$= 0.00025 / 0.00725$

$= 0.0345$ or 3.45%.

$= 0.0345 \times 100$

$= 3.45\%$.

The Probability that defective phone is manufactured at factory C is 3.45%.

Problem 3 : Feature Transformation & Kernels.

Desigigning Transformation.

1. Consider the 1-D data set as figure.

A. Yes, 1-D transformation to make points linearly separable is $f(x) = x^2$, which maps original 1D points to distinct positive and negative values, facilitating linear separation.

2. Still consider above 1-D dataset. Can you come up with a 2-D transformation that makes points linearly separatable!

A. Yes, for 1-D data set, a possible 2-D transformation could be map each point x to a 2-D point (x, x^2).

3. You may not always need to map a higher dimensional space to make the data linearly separatable, you may consider the 2-D data set as shown in fig 1(b).

A. Yes, for a 2-D data set, a possible 1-D transformation could be to project each point onto one of the axes. if points not linearly separatable they might be along the other.

4. Using ideas from the above Two datasets can you suggest a 2-D transformation of the data set as shown in Fig 1(c)

A. Yes, for data set fig 1 (c), a possible 2-D transformation could be to map each point (x, y) to a new point $(x^2 - y^2, 2xy)$. This transformation can make circularly distributed points linearly separable.

For kernal functions.

1. $K(x, z) = (xz + 1)^2$

Yes, the function $K(x, z) = (xz + 1)^2$ is valid kernel. It can be expressed as dot product in higher dimensional space specifically. If we let

$\phi(x)$ Symmetry: $K(x, z) = K(z, x)$

2. $K(x, z) = (xz - 1)^3$

A. No, the function $K(x,z) = (xz-1)^3$ is not a valid kernal. It can not express as a dot product in a higher dimensional space.

It does not satisfy the positivity property. because it hat negative values for both x and y.

Problem 4:

1. Consider the gometric distribution parameterized by ϕ

$P(y;\phi) = (1-\phi)^{y-1}\phi, y = 1, 2, 3 \ldots$

Show the geometric distribution of $b(y)$, η, $T(y)$ and $a(\eta)$

Solution:

It is an indeed family member of exponential.

the geometric distribution expressed as

$P(y;\phi) = (1-\phi)^{y-1}.\phi$

where $y = 1, 2, 3 \ldots$

$\therefore$ The geometric distribution is in the exponential family can be expressed as

$P(y;\phi) = b(y) * \exp(\eta * T(y) - a(\eta))$.

$\eta = \log(\phi / (1-\phi))$ $\therefore$ It's an natural Parameter

$T(y) = y$ $\therefore$ Sufficient statistic

$b(y) = 1$ $\therefore$ Base measure

$a(\eta) = -\log(1-\phi)$ $\therefore$ log Partition function

By substituting in the general form.

$$P(y;\phi) = (1-\phi)^{y-1} \cdot \phi = 1 * \exp(\log(\phi/(1-\phi)) * y - (-\log(1-\phi))).$$

Q. Given a training set $\{(x_n, y_n)\}_{n=1}^{N}$ and let the log-likehood of an example of an example be $\log P(y_n / x_n; w)$. rule by learning using a Glm model.

Solution

Let us assume of GLM model. The three assumptions are.

1. The response variable y follows as exponential family distribution

2. η is linearly related to predictor variable of x.

3. variable y is conditionally independent of other observation given x.

Probability mass function is given by with responses of y is

$$P(y;\theta) = (1-\theta)^{(y-1)}\,\theta$$

∵ θ is the probability of success in Bernoulli trial.

• exponential family of distribution can written in form as

$$P(y;\theta) = \exp(y\log(1-\theta)+\log(\theta))$$

The log-likelihood of training given

$$L(W) = \Sigma[i=1 \text{ to } N] \log P(y_i|x^i;W) = \Sigma[i=1 \text{ to } N]$$

$$y_i\log(1-\theta_i)+\log(\theta)$$

derivative with respect to W.

$$\nabla L(W) = \Sigma[i=1 \text{ to } N] (y_i - g(x_i^T W))x_i$$

∵ g is the inverse link function

The inverse link function is.

$$g(\eta) = \log(\theta/(1-\theta)) = \eta$$

∴ The ~~geome~~ stochastic gradient ascent rule for learning using a GLM model with the geometric y is given by.

$$W := W + \alpha(y_i - g(x_i^T W))x_i$$

∵ α is the learning rate.