# BLG527E Machine Learning Homework 1

Elif Özcan 504221508

October 23, 2022

## 1 Qestion 1

### 1.1 Evaluate the conditional probability mass function P(X | Z=1)

$$P(X \mid Z = 1) = \frac{P(X, Z = 1)}{P(Z = 1)}$$

$$P(X = 0 \mid Z = 1) = \frac{0.2}{0.3} = 0.66$$

$$P(X = 1 \mid Z = 1) = \frac{0.1}{0.3} = 0.33$$

### 1.2 Are random variables X and Z independent? Show your solution.

If two events are independent then their joint probability should be eaual to product of their marginal probabilities.

$$P(X, Z) = P(X) * P(Z)$$

... we should check for the the example

$$P(X = 1, Z = 1) = 0.1$$

$$P(X = 1) * P(Z = 1) = 0.5 * 0.3 = 0.15$$

$$P(X = 1) * P(Z = 1) \neq 0.5 * 0.3 = 0.15$$

$$P(X, Z) \neq P(X) * P(Z)$$

which means X and Z are not independent events.

### 1.3 Given that Y is known, are X and Z conditionally independent? Show your solution.

Conditionally independent events are defined as:

$$P(X \cap Z \mid Y) = P(X \mid Y) * P(Z \mid Y)$$

$$P(X = 1 \cap Z = 1 \mid Y = 0) = \frac{0.01}{0.33}$$

$$P(X \mid Y) * P(Z \mid Y) = \frac{0.13}{0.33} * \frac{0.09}{0.33}$$

$$P(X = 1 \cap Z = 1 \mid Y = 0) \neq P(X = 1 \mid Y = 0) * P(Z = 1 \mid Y = 0)$$

Given that Y is known, X and Z aren't conditionally independent.

| X | Y | g(X,Y) | P(X,Y \| Z=1) |
|---|---|---|---|
| 0 | -1 | 1 | 0.26 |
| 0 | 0 | 1 | 0.26 |
| 0 | 1 | 1 | 0.13 |
| 1 | -1 | 0 | 0.03 |
| 1 | 0 | 1 | 0.03 |
| 1 | 1 | 2 | 0.26 |

Table 1: E[(g(X,Y) | Z=1)]

| X | Y | g(X,Y) | P(X,Y \| Z=2) |
|---|---|---|---|
| 0 | -1 | 1 | 0.17 |
| 0 | 0 | 1 | 0.17 |
| 0 | 1 | 1 | 0.08 |
| 1 | -1 | 0 | 0.17 |
| 1 | 0 | 1 | 0.17 |
| 1 | 1 | 2 | 0.22 |

Table 2: E[(g(X,Y) | Z=2)]

### 1.4 Evaluate the conditional expectation of the following function g(X, Y | Z) = X × Y + 1|Z. (Note that,you need to evaluate the expectation for both values of Z.)

$$E[(g(X,Y) \mid Z=1)] = \sum(g(X,Y) * P(X,Y \mid Z=1))$$

$$E[(g(X,Y) \mid Z=1)] = \sum(g(X,Y) * P(X,Y \mid Z=1)) = 0.94$$

$$E[(g(X,Y) \mid Z=1)] = \sum(g(X,Y) * P(X,Y \mid Z=2))$$

$$E[(g(X,Y) \mid Z=1)] = \sum(g(X,Y) * P(X,Y \mid Z=2)) = 0.76$$

## 2 Question 2

### 2.1 Find the probability that a page visit results in product purchase.

In probability density function probability of a point is equal to zero. So if we were to accept only at p=1 the customer purchase a product then its probability will be equal to 0.

$$P(p=1) = \int_1^1 f_P(p)\,dp = 0$$

But if we accept p= [ 1/2 , 1] is where the purchase happens then its probability will be approximately 0.42.

$$P(1/2 \leq p \leq 1) = \int_{\frac{1}{2}}^1 f_P(p)\,dp = \frac{2*(e-\sqrt{e})}{5} = 0.42$$

### 2.2 Given that a page visit resulted in product purchase, find the conditional PDF of P.

Again in probability density function probability of a point is equal to zero.

$$P(p=1 \mid V=1) = 0$$

But if we accept p= [ 1/2 , 1] is where the purchase happens then its probability will be as follows,

$$P(1/2 \leq p \leq 1 \mid V = 1) = \int_{\frac{1}{2}}^{1} f_{P|V}(p \mid v)\, dp$$

But we don't have the $f_{P|V}(p \mid v)$ function. The parameter V represents visit and have 2 value 0,1. So we can write $f_{P|V}(p \mid v) = f_P(p) * v$ seeing if v=0 purchase can't happen and if v=1 probability of purchase is $f_P(p)$.

$$P(1/2 \leq p \leq 1 \mid V = 1) = \int_{\frac{1}{2}}^{1} f_{P|V}(p \mid v)\, dp = \int_{\frac{1}{2}}^{1} 0.4 * e^p * v\, dp = \frac{2 * (e - \sqrt{e}) * 1}{5} = 0.42$$

### 2.3 Given that the page visit resulted in product purchase, find the conditional probability of product purchase in the next page visit from the same customer.

Customers purchased the product according to the PDF independently so the first purchase don't affect to next purchase. The probability would be the same as above.

## 3 Question 3

Prove the followings where X and Y are random variables with joint distribution p(x, y).

$$E[X] = E_Y[E_X[X|Y]]$$

We know that $E[f(x)] = \sum p(x) * f(x)$ . Let $f(y) = E_X[X \mid Y = y]$ . Then,

$$E_Y[E_X[X|Y]] = E_Y[f(y)] = \sum_{y \in Y} p(y) * f(y)$$

$$E_Y[E_X[X|Y]] = \sum_{y \in Y} p(y) * E_X[X \mid Y = y]$$

$$E_Y[E_X[X|Y]] = \sum_{y \in Y} p(y) * \sum_{x \in X} p(x \mid y) * x$$

$$E_Y[E_X[X|Y]] = \sum_{x \in X} x * \sum p(x, y)$$

$$E_Y[E_X[X|Y]] = \sum_{x \in X} x * p(x)$$

$$E_Y[E_X[X|Y]] = E[X]$$

So for $var[X] = E_Y[var_X[X \mid Y]] + var_Y[E_X[X \mid Y]]$;

$$Var[X] = E[X^2] - E[X]^2$$

$$E[X^2] = E[E[X^2 \mid Y]] = E[Var[X \mid Y] + [E[X \mid Y]]^2]$$
$$E[X^2] - E[X]^2 = E[Var[X \mid Y] + [E[X \mid Y]]^2] - [E[E[X \mid Y]]]^2$$
$$E[X^2] - E[X]^2 = (E[Var[X \mid Y]) + ([E[X \mid Y]]^2] - [E[E[X \mid Y]]]^2)$$
$$E[X^2] - E[X]^2 = E_Y[var_X[X \mid Y]] + var_Y[E_X[X \mid Y]].$$

# 4 Question 4

## 4.1 Write down the two class discrimination function g(t) where t denotes temperature.

There are 2 classes which are sick and healthy or 1 and 0.

$$P(C \mid X) = \frac{P(X \mid C) * P(C)}{P(X)}$$

$$P(C \mid X) \propto P(X \mid C) * P(C)$$

$$g_i(t) = logP(t \mid C) + logP(C_i)$$

$$g_i(t) = \frac{-log2\pi}{2} - log\sigma_i - \frac{(t-\mu)^2}{2 * \sigma_i^2} + logP(C_i)$$

$$g_0(t) = \frac{-log2\pi}{2} - log1 - \frac{(t-36.5)^2}{2 * 1^2} + logP(C=0)$$

$$g_1(t) = \frac{-log2\pi}{2} - log1 - \frac{(t-39.2)^2}{2 * 1^2} + logP(C=1)$$

## 4.2 What is the decision threshold (in terms of temperature) in a two-class, two-action decision problem

Risk of action 1 and action follows as:

$$R(\alpha_1 \mid X) = 0 * P(C_1 \mid X) + \lambda_{12} * P(C_2 \mid X) = \lambda_{12} * (1 - P(C_1 \mid X))$$

$$R(\alpha_2 \mid X) = \lambda_{21} * P(C_1 \mid X) + 0 * P(C_2 \mid X)$$

where X represents temperature and C represents class.
If the cost of misclassifying the sick subjects is the same cost as the cost of misclassifying the healthy ones.

$$\lambda_{12} = \lambda_{21}$$

$$R(\alpha_1 \mid X) = \lambda_{12} * (1 - P(C_1 \mid X))$$

$$R(\alpha_2 \mid X) = \lambda_{21} * P(C_1 \mid X)$$

If the two misclassifications were equally costly, the decision threshold would be at 1/2. So Threshold will be $T = \frac{36.5 + 39.2}{2} = 37.85$
If the cost of misclassifying the sick subjects is three times the cost of misclassifying the healthy ones.

$$\lambda_{12} = 3 * \lambda_{21}$$

$$R(\alpha_1 \mid X) = 3 * \lambda_{21} * (1 - P(C_1 \mid X))$$

$$R(\alpha_2 \mid X) = \lambda_{21} * P(C_1 \mid X)$$

If $R(\alpha_1 \mid X) < R(\alpha_2 \mid X)$ then we choose $\alpha_1$.

$$3 * \lambda_{21} * (1 - P(C_1 \mid X)) < \lambda_{21} * P(C_1 \mid X)$$

$$3 * (1 - P(C_1 \mid X)) < P(C_1 \mid X)$$

$$\frac{3}{4} < P(C_1 \mid X)$$

So the decision threshold would be at 3/4.
If the cost of misclassifying the sick subjects is 18 times the cost of misclassifying the healthy ones.

$$\lambda_{12} = 18 * \lambda_{21}$$

$$R(\alpha_1 \mid X) = 18 * \lambda_{21} * (1 - P(C_1 \mid X))$$

$$R(\alpha_2 \mid X) = \lambda_{21} * P(C_1 \mid X)$$

If $R(\alpha_1 \mid X) < R(\alpha_2 \mid X)$ then we choose $\alpha_1$.

$$18 * \lambda_{21} * (1 - P(C_1 \mid X)) < \lambda_{21} * P(C_1 \mid X)$$
$$18 * (1 - P(C_1 \mid X)) < P(C_1 \mid X)$$
$$\frac{18}{19} < P(C_1 \mid X)$$

So the decision threshold would be at 18/19.
If the two misclassifications were equally costly, the decision threshold would be at 1/2 but because the cost of wrongly choosing C=1 is higher, we want to choose C=1 only when we are really certain, so the temperature shifts toward the class that incurs higher risk when misclassified.

### 4.3 Assume that the cost of misclassifying the sick subjects is 7 times higher than misclassifying the healthy subjects. What if we introduce another action, being indecisive, with the cost equal to the one-third of the cost of misclassifying healthy subjects. Give the temperature range for all three actions.

$$\lambda_{12} = 7 * \lambda_{21} = 21 * \lambda_3$$

For the sake of simplicity lets use number in place of cost values such as $\lambda_{12} = 21, \lambda_{21} = 3, \lambda_3 = 1$.

| Actions | sick | healthy |
|---|---|---|
| $\alpha_1$ | 0 | 21 |
| $\alpha_2$ | 3 | 0 |
| $\alpha_3$ | 1 | 1 |

$$R(\alpha_1 \mid X) = 21 * (1 - P(C_1 \mid X))$$
$$R(\alpha_2 \mid X) = 3 * (P(C_1 \mid X)$$
$$R(\alpha_3 \mid X) = 1$$

If $R(\alpha_1 \mid X) < 1$ then we choose $\alpha_1$.

$$21 * (1 - P(C_1 \mid X)) < 1$$
$$\frac{20}{21} < P(C_1 \mid X)$$

If $R(\alpha_2 \mid X) < 1$ then we choose $\alpha_2$.

$$3 * P(C_1 \mid X) < 1$$
$$P(C_1 \mid X) < \frac{1}{3}$$

If not either of them then we reject. $\alpha_1 \rightarrow (39.07, \infty)$, $\alpha_2 \rightarrow (-\infty, 37.40)$, $\alpha_3 \rightarrow (37.40, 39.07)$

# 5 Question 5

## 5.1 Derive the expected square error in terms of bias and variance. Estimate expected square error, bias, and variance for 1st and 2nd degree polynomial models.

$$y = Q_0 + Q_1 + \varepsilon = f(x)$$

$$E[(f(x) + \varepsilon - g(x))^2] = E[\varepsilon^2] + 2 * E[\varepsilon] * E[f(x) - g(x)] + E[(f(x) - g(x))^2]$$

$$E[\varepsilon] = 0$$

$$E[(f(x) + \varepsilon - g(x))^2] = E[\varepsilon^2] + E[(f(x) - g(x))^2]$$

In the above equation $E[\varepsilon^2]$ is the noise and $E[(f(x) - g(x))^2]$ is the squared error.

$$E[(f(x) - g(x))^2] = E[(f(x) - E[g(x)] + E[g(x)] - g(x))^2$$

$(f(x) - E[g(x)])^2$ is the square of bias and $E[(g(x) - E[g(x)])^2]$ is the variance.
For the 1st degree polynomial model variance will be lower than 2nd degree polynomial model and bias will be higer.

$$E[g(x)] = g'(x) = \frac{\sum g_i(x)}{M}$$

$$bias^2 = \frac{\sum [g'(x^t) - f(x^t)]^2}{N}$$

$$variance = \frac{\sum \sum [g'(x^t) - f(x^t)]^2}{N * M}$$