به نام خدا

دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر

پاسخ تمرین سری دوم شناسایی آماری الگو

استاد:

دکتر رحمتی

دانشجو:

حلیمه رحیمی

شماره دانشجویی:

۹۹۱۳۱۰۴۳

پاییز ۱۳۹۹

Assingment II

1. Do you think you may have COVID-19?

a. How worrying is the situation? Use Bayes' theorem to justify your answer.

(TPR + FNR is supposed to be equal to 1! However I kept the given values)

$P(Cov) = 0.004$ $TPR = Recall = 0.83$ $FPR = 0.17$ $FNR = 0.11$

$$P(Cov \mid Pos) = \frac{P(POS \mid COV)\, P(COV)}{P(Pos)} = \frac{P(POS \mid COV) P(COV)}{P(POS \mid Cov) P(Cov) + P(Pos \mid NoCov) P(NoCov)}$$

$$= \frac{0.83 \times 0.004}{0.83 \times 0.004 + 0.17 \times 0.996} = \frac{0.00332}{0.00332 + 0.16932} = \frac{0.00332}{0.17264} \simeq 0.0192$$

b. What is the probability that you actually have COVID-19?

$P(Cov) = 0.001$ $Accuracy = 0.73$ $FPR = 0.27$ $FNR = 0.19$

$TPR = 1 - FNR = 1 - 0.19 = 0.81$

$$P(Cov / Pos) = \frac{0.81 \times 0.001}{0.81 \times 0.001 + 0.27 \times 0.999} = \frac{0.00081}{0.27054} \simeq 0.003$$

c. How does your answer in the previous part change?

$P(Cov) = 0.05$ $$P(Cov \mid Pos) = \frac{0.81 \times 0.05}{0.81 \times 0.05 + 0.27 \times 0.95} = \frac{0.0405}{0.297} \simeq 0.136$$

d. How likely is your patient to actually have coronavirus when he tests negative?

(There was an issue here regarding calculating P(Cov). With the given values, P(Cov) became a negative number. So I considered P(Cov) = 0.025 instead of P(pos) = 0.025. I hope this is ok with you :))

$P(Pos|Cov) = 0.76$ $P(Neg|NoCov) = 0.68$ $P(Cov) = 0.025$

$P(Neg|Cov) = 1 - P(Pos|Cov) = 1 - 0.76 = 0.24$ $P(Pos|NoCov) = 0.32$

$$P(Cov|Neg) = \frac{P(Neg|Cov)P(Cov)}{P(Neg|Cov)P(Cov) + P(Neg|NoCov)P(NoCov)}$$

$$= \frac{0.24 \times 0.025}{0.24 \times 0.025 + 0.68 \times 0.975} = \frac{0.006}{0.669} \simeq 0.009$$

e. How about if the results show positive?

$$P(Cov|Pos) = \frac{P(Pos|Cov)P(Cov)}{P(Pos|Cov)P(Cov) + P(Pos|NoCov)P(NoCov)}$$

$$= \frac{0.76 \times 0.025}{0.76 \times 0.025 + 0.32 \times 0.975} = \frac{0.019}{0.331} \simeq 0.057$$

f. If your patient tests positive, is he more likely to actually have Coronavirus than not?

$P(NoCov|Pos) = 1 - P(Cov|Pos) = 1 - 0.057 = 0.943$

It's more likely to not have Coronavirus.

g. If your patient tests negative, is he more likely to not have Coronavirus than have it?

$P(NoCov|Neg) = 1 - P(Cov|Neg) = 1 - 0.009 = 0.991$

It's more likely to not have Coronavirus.

h. If a patient test positive, what is the probability of him actually has Coronavirus?

$P(Cov) = 0.01 \qquad P(Pos|Cov) = p \qquad P(Pos|NoCov) = 1-p$

$$P(Cov|Pos) = \frac{P(Pos|Cov)P(Cov)}{P(Pos|Cov)P(Cov) + P(Pos|NoCov)P(NoCov)}$$

$$= \frac{p \times 0.01}{p \times 0.01 + (1-p) \times 0.99} = \frac{p}{99-98p}$$

i. Let q be the probability that a patient actually has Coronavirus if the test is positive. For what numerical value of q do you consider the test as being reliable?

$$q = \frac{p}{99-98p}$$

(There is the same issue with this question as question 'a' which creates an issue in this part.)

The problem here is: If $q=1 \Rightarrow 1 = \frac{p}{99-98p} \Rightarrow 99-98p = p$

$\Rightarrow 99 = 99p \Rightarrow p = 1$ (There is no trade-off between precision and recall here.(!) while the model is the same and doesn't change.)

j. Based on parts h. and i., what should p be for you to consider the test as being reliable?

Assuming there were no issues with these questions, there would be a trade-off between precision and recall and because this is a matter of health, we need recall to be high.

Just for the sake of question let's consider $q=0.1$, then we would get $p \doteq 0.916$.

2. Making Decisions using Bayes Decision Rule

a. Sketch the contours of constant values for two class conditional densities.

The code for these questions can be found in

b. Sketch the decision boundary for Bayesian classifier and minimum distance classifier.

Since no prior probabilities were given, I considered them equal. Therefore $d_{12} = p(x|\omega_1) - p(x|\omega_2)$.

c. Generate 1000 samples for each class and estimate the classification error for each classifier.

Bayesian Classifier Error = 0.0155 MD Classifier Error = 0.0255

I considered ω_1 as positive (+) and ω_2 as negative (−).

Bayesian Classifier

	+	−	
+	985	15	Actual Class
−	16	984	
	Predicted Class		

Minimum Distance Classifier

	+	−	
+	998	2	Actual Class
−	49	951	
	Predicted Class		

d. Considering the following cost matrix, find the decision boundary for Bayes classifier and compare f_score for generated samples with Bayes classifier in part b.

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix} \Rightarrow d_{12} = p(x|w_1) - 3\,p(x|w_2)$$

Part b Classifier F_Score = 0.9845

$$F\_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Part d Classifier F_Score = 0.9808

	+	−	
+	971	29	Actual Class
−	991	9	
	Predicted Class		

$$\text{Classification Error} = \frac{FP + FN}{P + N}$$

$$= 0.019$$

e. Repeat part c. for 20 times, and report an average error for both classifiers.

Average Bayesian Classifier Error = 0.0143

Average Minimum Distance Classifier Error = 0.0268

f. In each case, sketch the Bayes decision boundary. If the classifier breaks down, explain.

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\|x-\mu_i\|^2}{2\sigma_i^2}\right) \qquad |\Sigma| = \sigma^2 \qquad \Sigma^{-1} = \frac{1}{\sigma^2} I$$

$$p(\omega_1|x) = p(\omega_2|x) \Rightarrow \frac{1}{\sigma_1} \exp\left(-\frac{\|x-\mu_1\|^2}{2\sigma_1^2}\right) p(\omega_1) = \frac{1}{\sigma_2} \exp\left(-\frac{\|x-\mu_2\|^2}{2\sigma_2^2}\right) p(\omega_2)$$

$$\Rightarrow d_{12} = \ln\frac{\sigma_2}{\sigma_1} - \frac{\|x-\mu_1\|^2}{2\sigma_1^2} + \frac{\|x-\mu_2\|^2}{2\sigma_2^2} + \ln\frac{p(\omega_1)}{p(\omega_2)}$$

If by breaking down we mean acting random, we're not having that here. However in picture 4 classifier doesn't work quite well.

If Σ was the same for all classes the classifier would have had a break down in picture 2 (But clearly Σs are not the same here, Picture 2 has equal mean for classes).

3. Automatic Saffron Cleaning Machine

a. Design three Bayesian classifiers using each of the feature sets. The data are assumed to have Gaussian distribution with the same covariance matrix $\Sigma = I$. For each of the classifiers, find the general form of the discriminant function.

Feature set 1 :

$\mu_1 = [E(x|w_1)] = \frac{1}{5}(-47.55 - 70.45 - 55.91 - 12.55 - 57.28) = \frac{-243.74}{5} = -48.748$

$\mu_2 = \frac{1}{5}(74.56 + 71.91 + 25.57 + 33.19 + 79.07) = \frac{284.3}{5} = 56.86$

$\mu_3 = \frac{1}{5}(44.35 + 27.83 + 47.32 + 54.10 + 56.82) = \frac{230.42}{5} = 46.084$

$\Sigma = I \Rightarrow p(x|w_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|x-\mu_i\|^2}{2}\right)$

$\Rightarrow g_i = -\frac{\|x-\mu_i\|^2}{2} + \ln p(w_i) \Rightarrow g_i = -\frac{1}{2}(\underset{\text{مشترک}}{x^Tx} - 2\mu_i^T x + \mu_i^T\mu_i) + \ln p(w_i)$

$g_i = \mu_i x - \frac{1}{2}\mu_i^T\mu_i + \ln p(w_i)$ $\qquad P(w_i)$ same for all.

Here μ_i is 1×1, So $\qquad g_i = \mu_i x - \frac{\mu_i^2}{2}$

$g_1 = \mu_1 x - \frac{\mu_1^2}{2} \qquad g_2 = \mu_2 x - \frac{\mu_2^2}{2} \qquad g_3 = \mu_3 x - \frac{\mu_3^2}{2}$

Feature Set 2:

$\mu_1 = \begin{bmatrix} E(x_1|w_1) \\ E(x_2|w_1) \end{bmatrix} = \begin{bmatrix} 11.6 \\ 0 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 2.6 \\ 6.4 \end{bmatrix} \quad \mu_3 = \begin{bmatrix} 22 \\ 1 \end{bmatrix}$

$g_i = \mu_i x - \frac{1}{2}\mu_i^T\mu_i \qquad g_i = \mu_{i1}x_1 + \mu_{i2}x_2 - \frac{1}{2}(\mu_{i1}^2 + \mu_{i2}^2)$

Feature set 3:

$\mu_1 = \begin{bmatrix} E(x_1|w_1) \\ E(x_2|w_1) \\ E(x_3|w_1) \end{bmatrix} = \begin{bmatrix} 0.742 \\ 0.622 \\ 0.696 \end{bmatrix}$

$$\mu_2 = \begin{bmatrix} 0.092 \\ 0.814 \\ 0.808 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 0.204 \\ 0.824 \\ 0.89 \end{bmatrix}$$

$$g_i = \mu_i x - \frac{1}{2} \mu_i^T \mu_i \qquad g_i = \mu_{i1} x_1 + \mu_{i2} x_2 + \mu_{i3} x_3 - \frac{1}{2} (\mu_{i1}^2 + \mu_{i2}^2 + \mu_{i3}^2)$$

b. A new saffron flower has arrived and an image has been recorded. The following features has been extracted from different parts of the image. Use the discriminant functions obtained in the previous part and predict their regions using each of the classifiers.

Feature Set 1:

#1 $g_1 = \mu_1 x - \frac{\mu_1^2}{2} = (-48.748 \times 54.73) - \frac{(-48.748)^2}{2} = -3856.161792$

$g_2 = (56.86 \times 54.73) - \frac{(56.86)^2}{2} = 1495.418$

$g_3 = (46.084 \times 54.73) - \frac{(46.084)^2}{2} = 1460.309792$

True Label : Stamen   Predicted Label : Stamen ✓

#2 $g_1 = -3678.231592$ $g_2 = 1287.879$ $g_3 = 1292.103192$

True Label : Stigma   Predicted Label : Stigma ✓

#3 $g_1 = 972.327608$ $g_2 = -4136.565$ $g_3 = -3104.310408$

True Label: Petal Predicted label: Petal ✓

#4 $g_1 = -3816.188432$ $g_2 = 1448.7928$ $g_3 = 1422.520912$

True Label: Stigma Predicted Label: Stamen ×

#5 $g_1 = 2053.558248$ $g_2 = -5397.7198$ $g_3 = -4126.453528$

True Label: Petal Predicted Label: Petal ✓

#6 $g_1 = -4534.246472$ $g_2 = 2286.3406$ $g_3 = 2101.338232$

True label: Stamen Predicted Label: Stamen ✓

#7 $g_1 = 1881.477808$ $g_2 = -5197.004$ $g_3 = -3963.777008$

True Label: Petal Predicted label: Petal ✓

#8 $g_1 = -3940.983312$ $g_2 = 1594.3544$ $g_3 = 1540.495952$

True label: Stamen Predicted Label: Stamen ✓

Feature Set 2:

Sample #	g_1	g_2	g_3	True Label	Predicted
#1	-67.28	-4.66	-239.5	Stamen	Stamen ✓
#2	2.32	17.34	-106.5	Stigma	Stamen ×

Sample#	g_1	g_2	g_3	True Label	Predicted	
#3	-32.48	35.14	-168.5	Petal	Stamen	×
#4	-67.28	-4.66	-239.5	Stigma	Stamen	×
#5	-67.28	8.14	-237.5	Petal	Stamen	×
#6	-67.28	84.94	-225.5	Stamen	Stamen	✓
#7	-67.28	20.94	-235.5	Petal	Stamen	×
#8	-67.28	14.54	-236.5	Stamen	Stamen	✓

Feature Set 3:

Sample#	g_1	g_2	g_3	True label	Predicted	
#1	0.444	0.715	0.697	Stamen	Stamen	✓
#2	0.398	0.71	0.683	Stigma	Stamen	×
#3	0.761	0.522	0.594	Petal	Petal	✓
#4	0.435	0.757	0.737	Stigma	Stamen	×
#5	0.895	0.687	0.758	Petal	Petal	✓

Sample #	g_1	g_2	g_3	True label	Predicted	
#6	0.547	0.807	0.804	Stamen	Stamen	✓
#7	0.822	0.614	0.673	Petal	Petal	✓
#8	0.457	0.732	0.713	Stamen	Stamen	✓

c. Compute a confusion matrix for each classifier.

Feature Set 1:

Actual Class \ Predicted Class	Petal	Stamen	Stigma
Petal	3	0	0
Stamen	0	3	0
Stigma	0	1	1

Feature set 2:

Actual Class \ Predicted Class	Petal	Stamen	Stigma
Petal	0	3	0
Stamen	0	3	0
Stigma	0	2	0

Feature Set 3:

Actual Class	Predicted Class: Petal	Stamen	Stigma
Petal	3	0	0
Stamen	0	3	0
Stigma	0	2	0

d. Calculate the Bayes error for each classifier.

Feature Set 1:

1 out of 8 test subjects have been classified wrong.

Error $= \frac{1}{8}$

Feature Set 2: Error $= \frac{5}{8}$

Feature Set 3: Error $= \frac{2}{8}$

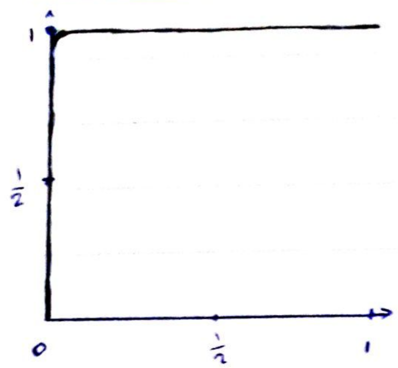
e. Draw an ROC curve to visualize the performance of each classification.

Feature Set 1:

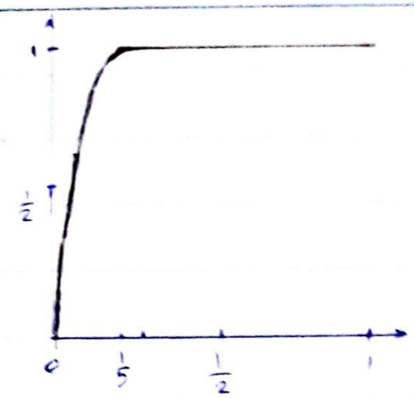
petal: $TPR = \frac{TP}{TP+FN} = \frac{3}{3} = 1$ $\quad FPR = \frac{FP}{FP+TN} = 0$

stamen: $TPR = \frac{3}{3+0} = 1$ $\quad FPR = \frac{1}{1+4} = \frac{1}{5}$

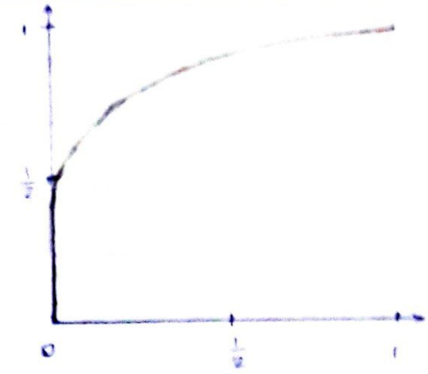
Stigma: $TPR = \frac{1}{1+1} = \frac{1}{2}$ $\quad FPR = 0$



(Petal)



(Stamen)



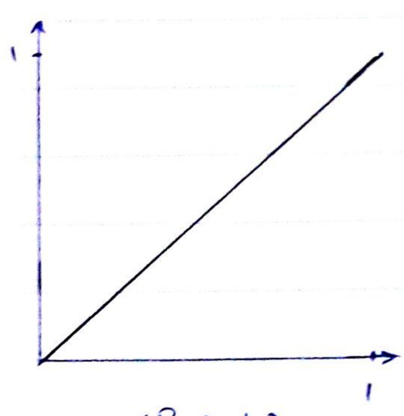
(Stigma)

Feature Set 2:

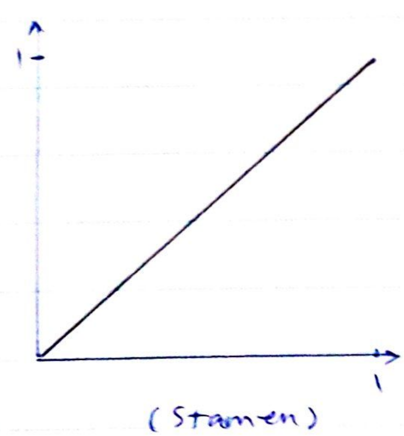
Petal: $TPR = 0$ $\quad FPR = 0$

Stamen: $TPR = \frac{3}{3} = 1$ $\quad FPR = \frac{5}{5+0} = 1$

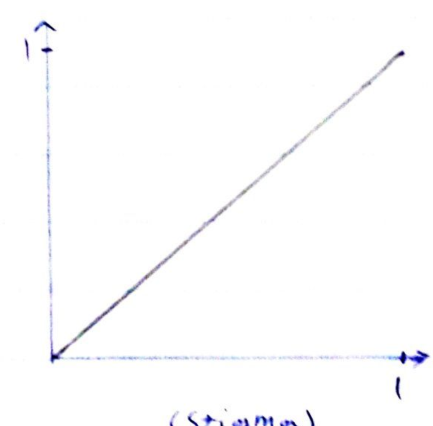
Stigma: $TPR = 0$ $\quad FPR = 0$



(Petal)



(Stamen)



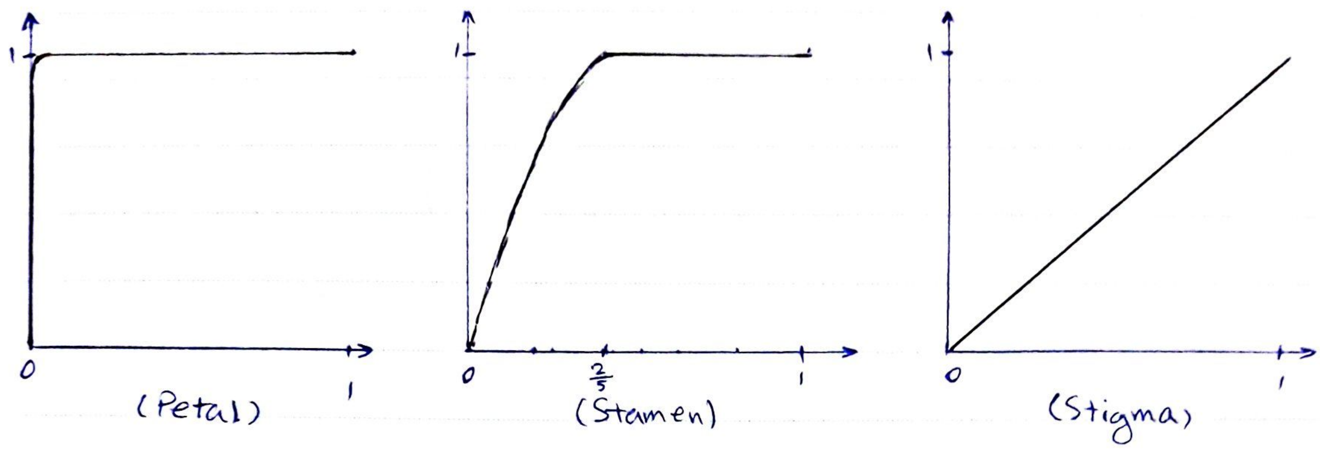
(Stigma)

Feature Set 3:

Petal: $TPR = \frac{3}{3+0} = 1$ $\quad FPR = 0$

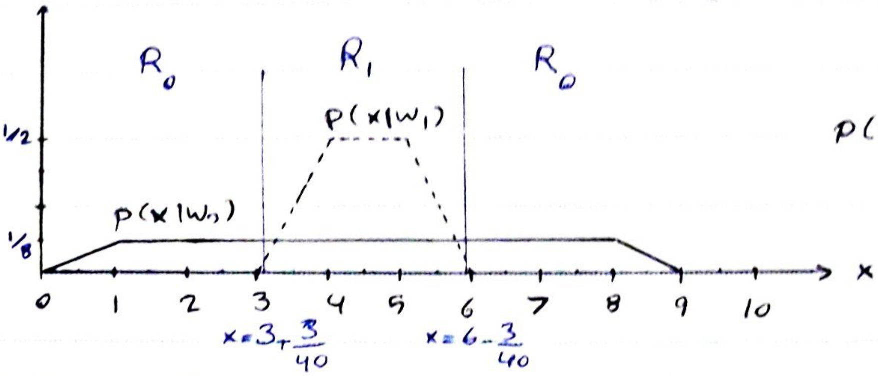
Stamen: $TPR = \frac{3}{3+0} = 1$ $\quad FPR = \frac{2}{2+3} = \frac{2}{5}$

Stigma : TPR = 0 FPR = 0



f. Repeat the previous parts with an MDC classifier. Display and compare the results.

4. Dealing with Prediction Error in Bayes Decision Rule



$p(w_0) = 0.3\,p(w_1)$

$P(x|w_1)$ has Trapezoidal Distribution.

$$p(x|w_0) = \begin{cases} \left(\frac{2}{9+8-0-1}\right)\left(\frac{x-0}{1-0}\right) = \frac{1}{8}x & 0 \leq x < 1 \\ \frac{2}{9+8-0-1} = \frac{1}{8} & 1 \leq x < 8 \\ \left(\frac{2}{9+8-0-1}\right)\left(\frac{9-x}{9-8}\right) = \frac{9-x}{8} & 8 \leq x \leq 9 \end{cases}$$

$$p(x|w_1) = \begin{cases} \frac{x-3}{2} & 3 \leq x < 4 \\ \frac{1}{2} & 4 \leq x < 5 \\ \frac{6-x}{2} & 5 \leq x \leq 6 \end{cases}$$

$p(w_0) = \frac{3}{13}$ $\quad$ $p(w_1) = \frac{10}{13}$

a. Determine and sketch the decision regions for a Bayes decision rule.

$$p(w_0|x) = p(w_1|x)$$

$$\frac{p(x|w_0)\,p(w_0)}{p(x)} = \frac{p(x|w_1)\,p(w_1)}{p(x)}$$

$$\frac{0.3\,p(x|w_0)\,\cancel{p(w_1)}}{\cancel{p(x)}} = \frac{p(x|w_1)\,\cancel{p(w_1)}}{\cancel{p(x)}}$$

Decision Boundaries

$$0.3\,p(x|w_0) = p(x|w_1)$$

$0 \leq x < 3$

$p(w_0|x) \propto 0.3\, p(x|w_0) > 0$ no boundaries here

$p(w_1|x) \propto p(x|w_1) = 0$ class w_0

$3 \leq x < 4$

$p(w_0|x) \propto 0.3\, p(x|w_0) = \frac{3}{80}$

$p(w_1|x) \propto p(x|w_1) = \frac{x-3}{2}$, $0 \leq \frac{x-3}{2} < \frac{1}{2}$

$\frac{3}{80} = \frac{x-3}{2} \Rightarrow x-3 = \frac{6}{80} \Rightarrow x = 3 + \frac{3}{40}$ decision boundary

$4 \leq x < 5$

$p(w_0|x) \propto 0.3\, p(x|w_0) = \frac{3}{80}$ no decision boundaries here

$p(w_1|x) \propto p(x|w_1) = \frac{1}{2}$ class w_1

$5 \leq x < 6$

$p(w_0|x) \propto 0.3\, p(x|w_0) = \frac{3}{80}$

$p(w_1|x) \propto p(x|w_1) = \frac{6-x}{2}$ $\frac{1}{2} \leq \frac{6-x}{2} < 0$

$\frac{3}{80} = \frac{6-x}{2} \Rightarrow x = 6 - \frac{3}{40}$ decision boundary

$6 \leq x \leq 9$

$p(w_0|x) \propto 0.3\, p(x|w_0) > 0$ no decision boundaries here

$p(w_1|x) \propto p(x|w_1) = 0$ class w_0

we have two Decision Boundaries : $x = 3 + \frac{3}{40}$, $x = 6 - \frac{3}{40}$

b. Compute the probability of error for this rule.

$$P(error) = P(\omega_1) \int_{R} p(x|\omega_1)\,dx + P(\omega_0) \int_{R_1} p(x|\omega_0)\,dx$$

$$P(error) = \frac{10}{13} \left(\int_{3}^{3+\frac{3}{40}} \frac{x-3}{2}\,dx + \int_{6-\frac{3}{40}}^{6} \frac{6-x}{2}\,dx \right) + \frac{3}{13} \int_{3+\frac{3}{40}}^{6-\frac{3}{40}} \frac{1}{8}\,dx$$

$$= \frac{10}{13} \left[\left(\frac{x^2}{4} - \frac{3}{2}x \right)_{3}^{3+\frac{3}{40}} + \left(3x - \frac{x^2}{4} \right)_{6-\frac{3}{40}}^{6} \right] + \frac{3}{13} \left(\frac{1}{8}x \right)_{3+\frac{3}{40}}^{6-\frac{3}{40}}$$

$$= \frac{10}{13} \left(\frac{9}{6400} + \frac{9}{6400} \right) + \frac{3}{13} \left(\frac{1}{8} \right) \left(3 - \frac{3}{20} \right) = \frac{9}{4160} + \frac{171}{2080} = \frac{351}{4160}$$

$$P(error) = 0.084375$$

c. Calculate the Bhattacharyya error bound.

$$\mathcal{E}_u = \sqrt{P(\omega_0)\,P(\omega_1)} \int \sqrt{p(x|\omega_0)\,p(x|\omega_1)}\,dx$$

$$p(x|\omega_0)\,p(x|\omega_1) = \begin{cases} \left(\frac{1}{8}x\right)(0) = 0 & 0 \leq x < 1 \\ \left(\frac{1}{8}\right)(0) = 0 & 1 \leq x < 3 \\ \left(\frac{1}{8}\right)\left(\frac{x-3}{2}\right) = \frac{x-3}{16} & 3 \leq x < 4 \\ \left(\frac{1}{8}\right)\left(\frac{1}{2}\right) = \frac{1}{16} & 4 \leq x < 5 \\ \left(\frac{1}{8}\right)\left(\frac{6-x}{2}\right) = \frac{6-x}{16} & 5 \leq x < 6 \\ \left(\frac{1}{8}\right)(0) = 0 & 6 \leq x < 8 \\ \left(\frac{1}{8}x\right)(0) = 0 & 8 \leq x \leq 9 \end{cases}$$

$$\mathcal{E}_u = \sqrt{\left(\frac{3}{13}\right)\left(\frac{10}{13}\right)} \left[\int_{3}^{4} \sqrt{\frac{x-3}{16}}\,dx + \int_{4}^{5} \sqrt{\frac{1}{16}}\,dx + \int_{5}^{6} \sqrt{\frac{6-x}{16}}\,dx \right]$$

$$\varepsilon_u = \frac{\sqrt{30}}{13*4}\left[\left(\frac{2}{3}(x-3)^{\frac{3}{2}}\right)_3^4 + (x)_4^5 + \left(-\frac{2}{3}(6-x)^{\frac{3}{2}}\right)_5^6\right]$$

$$= \frac{\sqrt{30}}{52}\left[\frac{2}{3}\left[(4-3)^{\frac{3}{2}} - (3-3)^{\frac{3}{2}}\right] + 1 - \frac{2}{3}\left[(6-6)^{\frac{3}{2}} - (6-5)^{\frac{3}{2}}\right]\right]$$

$$= \frac{\sqrt{30}}{52}\left[\frac{2}{3} + 1 + \frac{2}{3}\right] = \frac{\sqrt{30}}{52} * \frac{7}{3} \approx \frac{38.34}{156} \simeq 0.246$$

d. Calculate the Chernoff error bound.

$$\varepsilon_u = P^s(\omega_0)\,P^{1-s}(\omega_1)\int P^s(x|\omega_0)\,P^{1-s}(x|\omega_1)\,dx$$

$$= \left(\frac{3}{13}\right)^s\left(\frac{10}{13}\right)^{1-s}\left[\int_3^4 \left(\frac{1}{8}\right)^s\left(\frac{x-3}{2}\right)^{1-s}dx + \int_4^5 \left(\frac{1}{8}\right)^s\left(\frac{1}{2}\right)^{1-s}dx\right.$$

$$\left. + \int_5^6 \left(\frac{1}{8}\right)^s\left(\frac{6-x}{2}\right)^{1-s}dx\right] = \frac{(3^s)(10^{1-s})}{13}\left[\left(\frac{2^{s-1}(x-3)^{2-s}}{(2-s)\,8^s}\right)_3^4\right.$$

$$\left. + \left(\frac{2^{s-1}x}{8^s}\right)_4^5 + \left(-\frac{2^{s-1}(6-x)^{2-s}}{(2-s)\,8^s}\right)_5^6\right] = \frac{(3^s)(10^{1-s})}{13}\left[\frac{2^{s-1}}{(2-s)8^s}\left(\underbrace{(1-0)}_{1} - \underbrace{(0-1)}_{+1}\right)\right.$$

$$\left. + \frac{2^{s-1}}{8^s}\right] = \frac{(3^s)(10)^{1-s}}{13}\left[\frac{2^s}{(2-s)8^s} + \frac{2^{s-1}}{8^s}\right] = \left(\frac{(3^s)(10^{1-s})}{13}\right)\left(\frac{2^{s+1} - s2^{s-1}}{(2-s)8^s}\right)$$

If $s = \frac{1}{2} \Rightarrow \varepsilon_{Chernoff} = \varepsilon_{Bhattacharyya}$

I'm gonna simplify: $\varepsilon_u = \dfrac{5(s-4)3^s}{13(s-2)\,40^s}$, $0 \leq s \leq 1$

(The optimum s can be found by minimizing ε_u.)*

This means to minimize chernoff error bound in this s has to be 1.

*(Introduction to Statistical Pattern Recognition, Fukunaga, p98)

e. Determine the Neyman-Pearson classifier as well as its error.

Assume $\varepsilon_1 = 0.05$.

$$\varepsilon_1 = \int_{R_0} p(x|\omega_1)\,dx = 0.05$$

$$\varepsilon_1 = \int_3^{3+k} \frac{x-3}{2}\,dx + \int_{6-k}^{6} \frac{6-x}{2}\,dx = 0.05$$

$$\varepsilon_1 = \frac{1}{2}\left(\frac{x^2}{2} - 3x\right)\Big|_3^{3+k} + \frac{1}{2}\left(6x - \frac{x^2}{2}\right)\Big|_{6-k}^{6} = 0.05$$

$$\varepsilon_1 = \frac{(3+k)^2}{2} - 3(3+k) - \frac{9}{2} + 9 + 36 - \frac{36}{2} - 6(6-k) + \frac{(6-k)^2}{2} = 0.1$$

$$\varepsilon_1 = \cancel{\frac{9}{2}} + \frac{k^2}{2} + \cancel{\frac{6k}{2}} \cancel{-9} \cancel{-3k} \cancel{-\frac{9}{2}} \cancel{+9} \cancel{+36} \cancel{-\frac{36}{2}} \cancel{-36} \cancel{+6k} + \cancel{\frac{36}{2}} + \frac{k^2}{2} - \cancel{\frac{12k}{2}} = \frac{1}{10}$$

$$\varepsilon_1 = 2\left(\frac{k^2}{2}\right) = \frac{1}{10} \Rightarrow k^2 = \frac{1}{10} \Rightarrow k = \frac{1}{\sqrt{10}}$$

$$\underbrace{p(\omega_0|x) - p(\omega_1|x)}_{g_{01}(x)} \overset{\omega_0}{<} 3 + \frac{1}{\sqrt{10}} \qquad 3 + \frac{1}{\sqrt{10}} \overset{\omega_1}{<} g_{01}(x) \overset{\omega_1}{<} 6 - \frac{1}{\sqrt{10}} \qquad g_{01}(x) \overset{\omega_0}{>} 6 - \frac{1}{\sqrt{10}}$$

In other words: Decision Boundaries : $x = 3 + \frac{1}{\sqrt{10}}$, $x = 6 - \frac{1}{\sqrt{10}}$

ε_0 used to be equal to 0.35625 .

Now we have: $\varepsilon_0 = \displaystyle\int_{3+\frac{1}{\sqrt{10}}}^{6-\frac{1}{\sqrt{10}}} \frac{1}{8}\,dx = \frac{1}{8}x \Big|_{3+\frac{1}{\sqrt{10}}}^{6-\frac{1}{\sqrt{10}}}$

$$= \frac{1}{8}\left(6 - \frac{1}{\sqrt{10}} - 3 - \frac{1}{\sqrt{10}}\right) = \frac{1}{8}\left(3 - \frac{2}{\sqrt{10}}\right) \simeq 0.2959$$

$$P(\text{error}) = \left(\frac{10}{13}\right)\left(\frac{5}{100}\right) + \left(\frac{3}{13}\right)\left(\frac{3}{8} - \frac{1}{4\sqrt{10}}\right) \simeq 0.4529$$

ε_0 has decreased and $P(\text{error})$ has increased as expected.

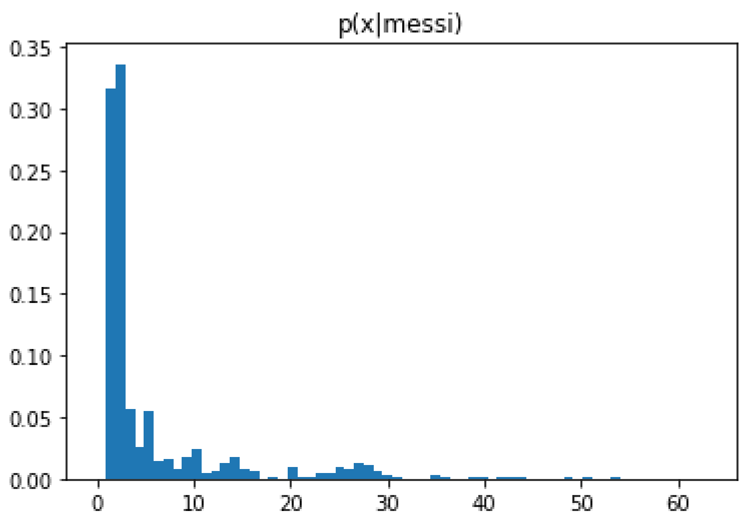
5.

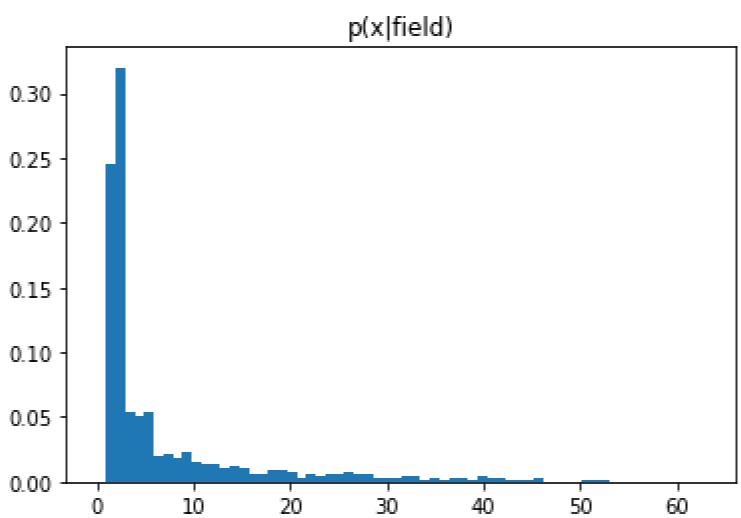
a. I counted the black and white pixels in each block and considered it white if the block had more white pixels. Then I calculated the prior by 'number of white blocks over number of blocks'

```
Prior of Messi:  0.35684313725490197  Prior of Field:  0.6431568627450981
```

b. I used 'normalized' form in histogram, so one can get the likelihoods like that.

I used the prior and likelihoods to calculate the posterior probabilities.

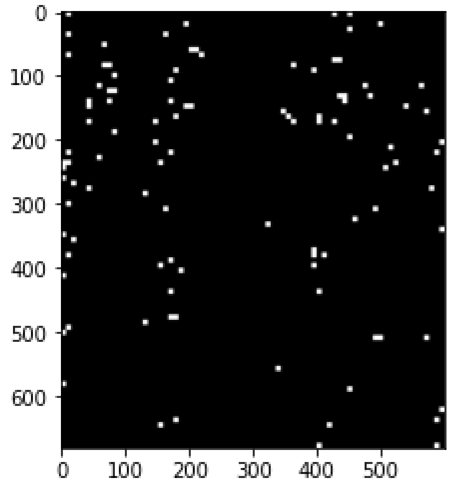





```
Posterior Probability of Messi 0.3601369436663557
Posterior Probability of Field 0.6557360722066604
```

c. I created a function get_features and used it to get the features.

d. and e. Unfortunately I don't know what has caused such an answer.




```
Error: 0.30633931276397
```

6. Maximum Likelihood Approach for Parameter Estimation

a. What is the likelihood function $L(\theta)$?

$$L(\theta) = \left(\frac{3\theta}{5}\right)^4 \left(\frac{2(1-\theta)}{5}\right)^2 \left(\frac{2\theta}{5}\right)^3 \left(\frac{3(1-\theta)}{5}\right)$$

$$= \left(\frac{(3)^4 (2)^3 (\theta^7)}{5^7}\right)\left(\frac{(2^2)(3)(1-\theta)^3}{5^3}\right) = \frac{6^5\ \theta^7\ (1-\theta)^3}{5^{10}}$$

b. Find the log likelihood function.

$$\ln L(\theta) = 5\ln 6 - 10\ln 5 + 7\ln\theta + 3\ln(1-\theta)$$

c. Using one of the above functions, determine the maximum likelihood estimate of θ.

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{7}{\theta} - \frac{3}{1-\theta} = 0 \Rightarrow 7(1-\theta) - 3\theta = 0 \Rightarrow 7 - 10\theta = 0 \Rightarrow \theta = \frac{7}{10}$$

d. Find the maximum likelihood estimate of the parameter θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{(m-1)!} \left(\frac{1}{\theta}\right)^m x^{m-1} e^{-\frac{x}{\theta}}$$

$$\ln L(\theta) = \sum_{i=1}^{n}\left(\ln\frac{1}{(m-1)!} - m\ln\theta + (m-1)\ln x_i - \frac{x_i}{\theta}\right)$$

$$= n\ln\frac{1}{(m-1)!} - nm\ln\theta + n(m-1)\sum_{i=1}^{n}\ln x_i - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{-nm}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0 \rightarrow -nm\theta + \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{nm}$$

e. Find $\hat{\theta}_{ML}$. $L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!}$

$$\ln L(\theta) = \sum_{i=1}^{n} (-\theta + y_i \ln\theta - \ln y_i!)$$

$$= -n\theta + \ln\theta \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \ln y_i!$$

$$\frac{\partial \ln L(\theta)}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} y_i}{\theta} = 0$$

$$\Rightarrow -n\theta + \sum_{i=1}^{n} y_i = 0 \Rightarrow \theta = \frac{\sum_{i=1}^{n} y_i}{n}$$

f. Prove that $\hat{\theta}_{ML}$ is unbiased for θ and find its variance.

$$E(\hat{\theta}) = E\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{n\theta}{n} = \theta$$

$$\mathrm{var}(\hat{\theta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(y_i) = \frac{n\theta}{n^2} = \frac{\theta}{n}$$

g. Assuming the gamma distribution as the prior distribution for θ, compute the posterior distribution and posterior mean of θ. Compare ML and MAP estimates.

$$P(\theta \mid D) \propto P(D\mid\theta)P(\theta) = \left(\prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!}\right)\left(\lambda e^{(-\lambda\theta)}\right)$$

$$\ln P(D\mid\theta)P(\theta) = -n\theta + \ln\theta\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\ln y_i! + \ln\lambda - \lambda\theta$$

$$\frac{\partial \ln P(D\mid\theta)P(\theta)}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} y_i}{\theta} - \lambda = 0 \Rightarrow -n\theta - \lambda\theta + \sum_{i=1}^{n} y_i = 0$$

$$\Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} y_i}{n+\lambda}$$

As far as I have learned we use mode in MAP, that's why I estimated θ the way I did in the page before.

On the other hand posterior mean of θ is an estimation for θ but we use mean of the posterior distribution.

meaning $\bar{\theta} = \int \theta \, p(\theta \mid D) d\theta = \dfrac{\int \theta \, L(\theta) p(\theta) d\theta}{\int L(\theta) p(\theta) d\theta}$.

Unfortunately I couldn't estimate this.

When it comes to comparing ML and MAP, Since we use prior probability of θ, there would be a shift in distribution and therefore we would get a biased estimate.

We could say ML is a special case of MAP where $p(\theta)$ is uniform.

h. Show that $T(y) = \sum_{n=1}^{N} y_n$ is a sufficient statistic for θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta} \theta^{y_i}}{y_i!} = (e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i})(\frac{1}{\prod_{i=1}^{n} y_i!})$$

Since we can write $P(D \mid \theta)$ as $g(S, \theta) h(D)$ then Yes, $T(y)$ is a sufficient statistic for θ.

7. Here's Why Hitler Hated MLE So Bad

a. Assuming n tanks are in the battlefield and k have been observed, find the likelihood of the highest observed serial number being m.

This is as if we have observed k out of n tanks so in general $\binom{n}{k}$, and up until $k-1$, we haven't observed the m we are considering $(m = M)$. Therefore:

$$P(m|n,k) = \frac{\binom{m-1}{k-1}}{\binom{n}{k}}$$

b. Find the posterior $P(n|m,k)$, considering a flat prior between the number of spotted tanks k and some maximum ~~number~~ Ω such that $p(n) = \begin{cases} 1/\Omega & \text{for } 1 \leq n < \Omega \\ 0 & \text{otherwise} \end{cases}$. Take the limit $\Omega \to \infty$.

$$p(n|m,k) = \frac{p(m|n,k)\, p(n|k)}{p(m|k)}$$

$$p(m|k) = p(m|k) \sum_{n=0}^{\infty} p(n|m,k) = p(m|k) \frac{\sum p(m|n,k)\, p(n|k)}{p(m|k)}$$

$$= \sum_{n=0}^{\infty} p(m|n,k)\, p(n|k)$$

$$\Rightarrow p(n|m,k) = \frac{p(m|n,k)\, p(n|k)}{\sum_{n=0}^{\infty} p(m|n,k)\, p(n|k)}$$

we know that we have observed m and $n < \Omega$.

$$\Rightarrow p(n|m,k) = \frac{p(m|n,k)}{\sum_{n=m}^{\Omega-1} p(m|n,k)} = \frac{\cancel{\binom{m-1}{k-1}} \binom{n}{k}^{-1}}{\cancel{\binom{m-1}{k-1}} \sum_{n=m}^{\Omega-1} \binom{n}{k}^{-1}} =$$

↳ given hint

$$\frac{\binom{n}{k}^{-1}}{\frac{1}{k-1}\left(\frac{(m-k)!}{(m-1)!}-\frac{(\Omega-k)!}{(\Omega-1)!}\right)}=\frac{\binom{n}{k}^{-1}}{\frac{k}{k-1}\left(\frac{1}{\binom{m-1}{k-1}}-\frac{1}{\binom{\Omega-1}{k-1}}\right)}$$ for $k \geq 2$

$$\binom{m-1}{k-1}=\frac{(m-1)!}{(m-k)!(k-1)!}$$

$$\lim_{\Omega\to\infty} p(n|m,k) = \lim_{\Omega\to\infty}\frac{\binom{n}{k}^{-1}}{\frac{k}{k-1}\left(\frac{1}{\binom{m-1}{k-1}}-\cancel{\frac{1}{\binom{\Omega-1}{k-1}}}\right)}=\frac{\binom{n}{k}^{-1}}{\frac{k}{k-1}\left(\frac{1}{\binom{m-1}{k-1}}\right)}=\frac{k-1\binom{m-1}{k-1}}{k\binom{n}{k}}$$

$\to \infty$

goes to zero

c. Find the posterior mean.

$$E(p(n|m,k))=\sum_{n=m}^{\infty} n\,\frac{k-1\binom{m-1}{k-1}}{k\binom{n}{k}}=\frac{k-1}{k}\binom{m-1}{k-1}\sum_{n=m}^{\infty}\frac{n}{\binom{n}{k}}=$$

$$\binom{n}{k}=\frac{n(n-1)!}{k!(n-k)!}$$

$$\frac{k-1}{k}\binom{m-1}{k-1}\sum_{n=m}^{\infty}\frac{k}{\binom{n-1}{k-1}}=\frac{\cancel{k}(k-1)}{\cancel{k}}\binom{m-1}{k-1}\sum_{n=m}^{\infty}\frac{1}{\binom{n-1}{k-1}}=$$

$$(k-1)\binom{m-1}{k-1}\left(\frac{k-1}{k-2}\right)\left(\frac{1}{\binom{m-2}{k-2}}-\frac{1}{\binom{\infty}{k-2}}\right)=\left(\frac{\cancel{(k-1)}(m-1)(m-2)!}{\cancel{(k-1)}(k-2)!(m-k)!}\right)\left(\frac{k-1}{k-2}\right)\left(\frac{1}{\binom{m-2}{k-2}}\right.$$

$$\left.-\frac{1}{\binom{\infty}{k-2}}\right)=\frac{(k-1)(m-1)}{k-2}\binom{m-2}{k-2}\left(\frac{1}{\binom{m-2}{k-2}}-\frac{1}{\binom{\infty}{k-2}}\right)=\frac{(k-1)(m-1)}{k-2}$$

$\to$ goes towards 0

d. Calculate the minimum number of observed tanks in order for the posterior mean to be finite.

k has to be equal to or more them 3. ($k \geq 3$)

I saw in the above equation any number ≥ 3 gives a finite number. I calculated for 2: $p(n|m,k=2)=\frac{2-1\binom{m-1}{2-1}}{2\binom{n}{2}}$

$$= \frac{m-1}{2\binom{n}{2}} = \frac{m-1}{\frac{n!}{(n-2)!}} = \frac{m-1}{\frac{n(n-1)(n-2)!}{(n-2)!}} = \frac{m-1}{n(n-1)}$$

$$E(p(n|m,k=2)) = \sum_{n=m}^{\infty} \cancel{n} \frac{m-1}{\cancel{n}(n-1)} = m-1 \sum_{n=m}^{\infty} \left(\frac{1}{n-1}\right) = \infty$$

e. Given $k=25$, $m=200$, and $\Omega=10000$, plot the posterior for the number n of tanks.

Around 300 probability gets to be about zero.

9. Some Explanatory Questions

a) In a binary classification problem, under what circumstances a Bayes classifier and a minimum distance classifier obtain exactly the same results?

When the covariance matrix of each class is equal to Identity matrix.

b. Does a minimum distance classifier have a training phase? What about a minimum-error classifier? Explain.

In minimum distance classifier we need the mean of each class therefore the training phase includes finding these parameters. Minimum-error classifier also has a training phase in which we learn about distribution parameters.

c. Assume a two-class 1D classification problem with the Gaussian distributions $p(x|\omega_1) \sim N(-1,1)$ and $p(x|\omega_2) \sim N(4,1)$, where the probabilities are equal. You are free to choose any classification method you would like, and you are given an infinitely large dataset. What would be the best error you can achieve on the test set, and why?

The best error would be Bayes Error and given the fact that the variance and probabilities are the same for both classes, I'm saying an MDC classifier would give the same amount of error.

In this case there would be a decision boundary at $x = 1.5$.

Since we have an infinitey large dataset, most of the data would be around the given μ. Since $\int_{R_1} p(x|\omega_0)\,dx = F(x > 1.5)$ and $R_1\{1.5, +\infty\}$

$P\{x > 1.5\} \propto P(z > 2.5) = 0.0062$, $\int_{R_0} p(x|\omega_1)\,dx = F(x < 1.5)$ and

$P\{x < 1.5\} \propto P(z < -2.5) = 0.0062$ So

$$P(Error) = p(\omega_0)\int_{R_1} p(x|\omega_0)\,dx + p(\omega_1)\int_{R_0} p(x|\omega_1)\,dx = \frac{1}{2}(0.0062 + 0.0062)$$

$= 0.0062$

d. Discuss whether it is possible to plot ROC curve for a classification problem with more than two catogaries? If yes, how? And if no, why?

Yes but you gotta do it one vs. all. This indicates the number of ROC curve plots would be equal to the number of classes. For each curve we consider one of the classes as positive class and the rest as negative, then we compute TPR and FPR as we did with only two classes being positive and negative. Here we are just adding FP, FN and TN through the rest of the classes (only the FP, FN and TN related to positive class).

e. Is it possible to apply the Bayesian Decision Rule in a regression problem? If yes, explain how. If no, explain why.

I did some search on this and what I found was "Bayesian Linear regression". As I understood we use bayesian inference here. So we don't really have the type of Decision Rule we had before. As I understand Bayesian Decision Rule is capable of finding boundaries that minimize the error, impicitly what happens in "Bayesian Linear Regression" is the same but we are not getting boundaries. Based on what I have said I consider the answer to be No; because we don't predict a number via Bayesiam Decision Rule.

f. Is the result of the Bayes decision rule unique? Explain.

Yes, since we are getting the minimum error and bayes error is the least error one can get.

g. When does MLE estimatien lead to a better result than MAP estimation?

When we are computing $p(\theta)$ from small amount of data. If the amount of data we are using in order to find the distribution of data is not big enough, it's possible that $p(\theta)$ would have a negative effect.

h. What is penalized MLE? When is it better to apply penalized MLE instead of normal MLE?

In Penalized MLE we consider model complexity. When we have a mixture of models, we're not gonna have an estimation that considers all the ups and downs of the probability in normal MLE method. However PMLE takes these in account by penalizing MLE In PMLE we gotta maximize log(likelihood) - p. For p there are several methods.