STAT 391

Final Exam

2:30 – 4:20 pm on June 8, 2018

mmp@stat.washington.edu

Student Name:____

You are allowed 8 pages of notes
Write your name clearly on each of the notes sheets; you will be asked to hand them in with the exam, but they will not be graded and will be returned to you in the statistics office starting on Monday. If you use additional paper (from us) you must write your name on it.

No electronic devices of any kind are allowed during the exam.

Any fact that was proved in the lectures or in the notes can be used without proof.

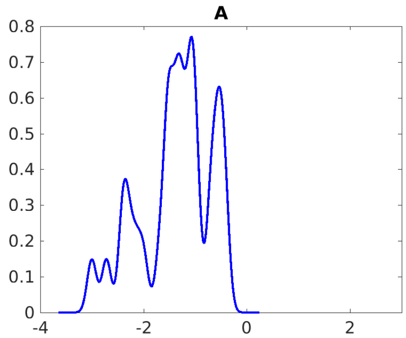
Do Well!

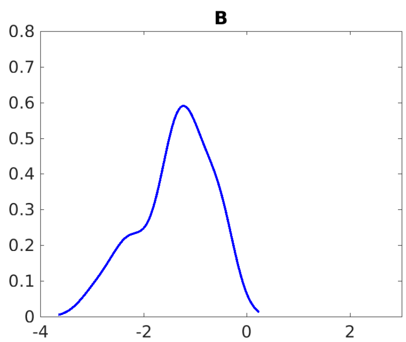
Prob.1 KDE _______	of 4
Prob.2 Poisson _______	of 3
Prob.3 Classifiers _______	of 8
Prob.4 BIC _______	of 14
Prob.5 Regression _______	of 7
TOTAL: _________	of 36

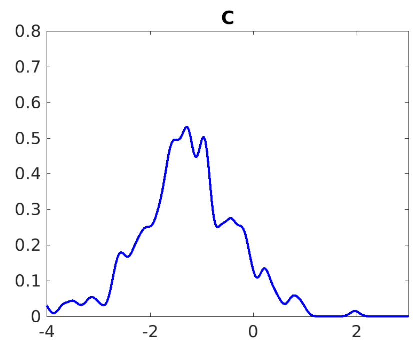
(4 points) **Problem 1 – KDE**

Plots **A, B, C** show Kernel Density Estimators, all using the Gaussian kernel with widths h_A, h_B, h_C.

1.1 Mark the true statements below each of **B,C**.







- ☐ Same data, $h_B > h_A$
- ☐ Same data, $h_B < h_A$
- ☐ Less data, $h_B < h_A$
- ☐ Less data, $h_B = h_A$

- ☐ Same data, $h_C > h_A$
- ☐ Less data, $h_C > h_A$
- ☐ More data, $h_C = h_A$
- ☐ Less data, $h_C = h_A$

1.2 Mark the correct answers below.

The estimator **A** has more BIAS than the estimator **B**.

TRUE FALSE

The estimator **C** has more VARIANCE than the estimator **A**.

TRUE FALSE

(3 points) **Problem 2 – ML estimation of Poisson distribution** *Show your work*

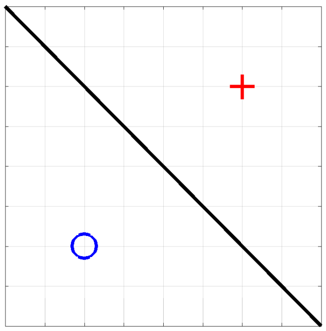
Data are sampled from a Poisson distribution $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x \in \{0, 1, 2, \ldots\}$. By mistake, zeros and ones were counted together and you are given the following data

Value x	0 or 1	2	3	4	
# observed	8	6	4	2	Total $n = 20$

Write the expression of the log-likelihood l of the data as a function of λ. *Literal OR numeric expression.*

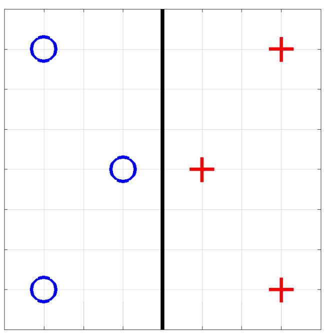
(7.5 points) **Problem 3 – Decision regions**

Below are examples of data and decision boundaries.



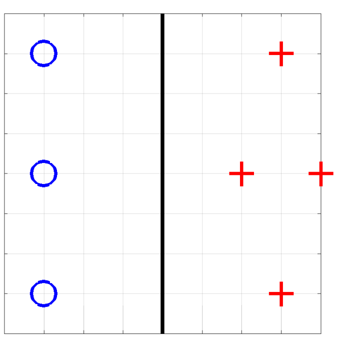
Mark all the classification methods below that could have produced this decision boundary.

- ☐ 1-nearest neighbor classifier
- ☐ Logistic regression
- ☐ Linear discriminant
- ☐ Quadratic discriminant



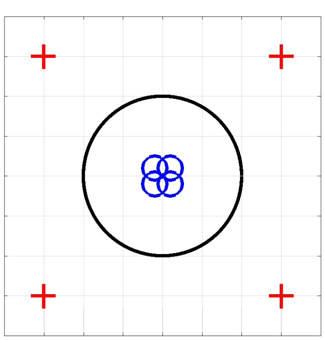
Mark all the classification methods below that could have produced this decision boundary.

- ☐ 1-nearest neighbor classifier
- ☐ Logistic regression
- ☐ Linear discriminant
- ☐ Quadratic discriminant



Mark all the classification methods below that could have produced this decision boundary.

- ☐ 1-nearest neighbor classifier
- ☐ Linear discriminant
- ☐ Quadratic discriminant



Mark all the classification methods below that could have produced this decision boundary.

- ☐ 1-nearest neighbor classifier
- ☐ Logistic regression
- ☐ Linear discriminant
- ☐ Quadratic discriminant

(14 points) **Problem 4 – Model selection by BIC** *Show your work*

For this problem, expressions of the form $2^{\pm m/k}$ are considered simple. We provide a table with numerical values for some of these powers, but you don't need to replace these values in your answers.

The data is $\mathcal{D} = \{x_1, \ldots x_n\}$ with $x_i \in [0, 1]$. We want to compare the following classes of densities on the real line:

$$\mathcal{M}_0 = \{\text{uniform}[0, 1]\},$$

$$\mathcal{M}_1 = \{\text{uniform}[0, b],\ b > 0\},$$

$$\mathcal{M}_2 = \{\text{uniform}[a, b],\ 0 \le a < b\}.$$

BIC(model class)=log-likelihood − #parameters×(log n)/2

Model	d	log-likelihood l	BIC(model)
$\mathcal{M}_0$		l_0	BIC($\mathcal{M}_0$) =
$\mathcal{M}_1$		l_1	BIC($\mathcal{M}_1$) =
$\mathcal{M}_2$		l_2	BIC($\mathcal{M}_2$) =

4.1 Fill in the column d with the number of parameters estimated from data for each of the models. *No need to show work for this question.*

4.2 For models $\mathcal{M}_{1,2}$, give the expression of the Maximum Likelihood estimates $b_{1,2}^{ML}, a_2^{ML}$. Show that $b_1^{ML} = b_2^{ML}$.

4.3 For simplicity, let us leave out the ^{ML} superscript for the remainder of this problem, and denote the estimates from **4.2** by a, b, remembering what they represent. Fill in the column "log-likelihood" with the expression of the log-likelihood l as a function of the ML parameters estimated for each of the three models $\mathcal{M}_{0,1,2}$ (e.g $l_2(a, b) = \ldots$). *Show your work here as needed.*

4.4 Fill in the column "BIC(model)" with the expression of the BIC criterion corresponding to each of the three models. *No need to show work for this question.*

4.5 Let $n = 16$ and let all logarithms above be base 2. For what values of b is $\mathcal{M}_0$ chosen over $\mathcal{M}_1$?

k	1	2	3	4	5	6	7	8	9	10
2^{-k}	0.5	0.25	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001
$2^{1/k}$	2	1.41	1.26	1.19	1.15	1.12	1.10	1.09	1.08	1.07
$2^{-1/k}$	0.5	0.71	0.79	0.84	0.87	0.89	0.90	0.92	0.93	0.93

4.6 Let $n = 16$ and let all logarithms above be base 2. For what values of a, b is $\mathcal{M}_2$ chosen over $\mathcal{M}_1$? *Find a simple relationship between* a, b

4.7 Suppose that the $n = 16$ samples are drawn from $\mathcal{M}_0$. What is the probability that $\mathcal{M}_0$ is preferred to $\mathcal{M}_1$ by the BIC criterion?

(7 points) **Problem 5 – Linear regression with input noise** *Show your work*

We are given a data set $\mathcal{D} = \{(x_1, y_1), \ldots (x_n, y_n)\}$ with $x_i, y_i \in \mathbb{R}$. Suppose the data were generated by

$$y_i \;=\; \beta(x_i + \epsilon_i), \text{ with } \epsilon_i \sim N(0, \sigma^2), \text{ i.i.d.} \tag{1}$$

The parameters β, σ^2 are unknown.

5.1 Write the likelihood of *of one data point* y_i given x_i under the model (1).

5.2 Write the expression of the log-likelihood $l(\beta, \sigma^2)$ for the whole data set $\mathcal{D}$.

5.3 Now find the ML estimate β^{ML} by maximimizing $l(\beta, \sigma^2)$ with respect to β.

5.4 Does the β^{ML} estimation problem have sufficient statistics? What are they?

5.5 What is the expectation of β^{ML}? *Remember to show your work*

[extra space on next page]

[extra space for anything]

Have a nice summer!