Midterm exam supplement. Closed book and calculators not allowed. Answers may include e^2, $\sqrt{\ }$, etc. but simplfy when possible.

Exam written by Paul Horton ©2019.

Problem 1.
Consider four classifiers simple classifiers:
Logistic Regression (LR), k Nearest Neighbors (kNN), Naïve Bayes (NB),
and Decision Tree (DT).

Which classifers fit the following statements?

(example of how to answer)
1- Is a classifier. LR, NN, NB, DT

1a Recursively partitions the data. ____________________

1b Has linear decision boundaries. ____________________

1c Is relatively robust to "the curse of dimensionality". ____________________

1d If given infinite data will converge to optimal classifier. ____________________

Problem 2.
Let $\mathrm{Y}(x)$ be a mixture model of two normal distributions $\mathrm{N}_1(\mu_1, \sigma_1^2)$, and $\mathrm{N}_2(\mu_2, \sigma_2^2)$. With mixture model coefficients (weights) of w_1 and w_2 respectively
$(w_1 + w_2 = 1)$.

2a Write down the probability density function for $\mathrm{Y}(x)$. (If you don't remember perfectly, write down the most important part)

2b What is the derivative of the natural logarithm of Y(x) with respect to μ_1? In other words, what is?

$$\frac{\partial}{\partial \mu_1} \; ln \, \mathrm{Y}(x)$$

Problem 3.
Consider a random variable X following a Bernoulli distribution over {a,b} with an unknown probability of a, which we denote as P_a. We assume a uniform distribution over P_a. In other words: $p[P_a] = 1, \;\; P_a \in [0,1]$. Sampling from X we observe the **S**= baa. For convenience we use F, F_a, F_b to denote the length of this sequence and the number of a's and b's it contains.
So for **S**: $F = 3, F_a = 2, F_b = 1$.

3a. What is likelihood of Pa given **S**, i.e. $P[\mathbf{S}|P_a]$?

3b. Recalling that we are using a uniform prior for P_a, what is the posterior probability distribution of Pa after seeing **S**?

3c. Given we have seen **S**, what is the probability that the next letter will be a?

For reference, Beta integral:

$$\int_0^1 dP_a \, P_a{}^{F_a}(1-P_a)^{F_b} = \frac{\Gamma(F_a+1)\Gamma(F_b+1)}{\Gamma(F_a+F_b+2)} \qquad = \frac{F_a!F_b!}{F_a+F_b+1!}, \text{for non-negative integers } F_a, F_b$$

Problem 4.
Let $Y \sim U(0,1)$ denote the uniformly distribution over [0,1], with probability density function:

$$P[y = x] = dx; \;\; 0 \leq x \leq 1.$$

Further let M_k denote the random variable obtained by taking the minimum of k independent samples from Y.
So M_2 is the minimum of two samples, etc.

4a. What is the probability density function of M_2?

4b. More generally, what is the probability density function of M_k?

Problem 5.
An basket contains k balls, of which b are black. One ball is drawn from the basket and then replaced. This is done n times.

Let n_b denote the number of times the ball drawn is black.

Question:

In terms of the variables k, b and n;

5a. What is the probability distribution of n_b?

5b. What is the mean, variance, and standard deviation of this probability distribution?

5c. What is the mean, variance, and standard deviation of n_b for the specific cases of: n=5, b=2, k=10; and n=225, b=2, k=10?

Problem 6.
There are five baskets, $u_0 \ldots u_4$, each contains 4 balls. Of the four balls they contain, u_0 has no black balls, u_1 has one black ball, ..., finally u_4 has only black balls (4/4).
First, a basket u is picked at random (with $u_0 \ldots u_4$ all having an equal chance).

Then 3 balls are randomly drawn from basket u, one at a time, replacing the ball each time.
In other words 3 balls are sampled from u with replacement.
Let b denote the number of the 3 drawn balls which are black.

Question:
In the case where $b = 2$, what is the probability that the next ball drawn will be black?
Show your work.

Exam, solutions by Paul Horton ©2019.