CMPS 242: Machine Learning: Fall 2016: HW 1 (Update 1)

Due: 11th October 2016

General Instructions (please review carefully):

- The assignment is to be *preferably* attempted in groups of two. If you choose to not work with a partner, points will **not** be automatically deducted from your score for this homework.
- Each group needs to submit only one set of solutions. However, each group member should completely understand the group's solutions. It is **strongly** recommended that each student work on all of the problems individually before integrating their solutions into the group consensus. Dividing up the problems so each student does just a subset of them is the wrong approach.
- LaTeX is preferred, but neatly handwritten solutions will also be accepted. All solutions need to be handed over in the class before the beginning of the lectures.
- The names of the group members, and their UCSC ID (@ucsc.edu email address) should prominently be written on the upper left corner of the first page.
- Multiple sheets should be stapled together in the upper **left** corner.
- Solutions to the problems should be clearly labeled with the problem number.
- Although no points are given for neatness, illegible and/or poorly organized solutions can be penalized at the TA discretion.
- Clearly acknowledge sources (web, people, books, etc.), and mention if you discussed the problems with other students or groups. In all cases, the course policy on collaboration applies, and you should refrain from getting direct answers from anybody or any source. If in doubt, please ask the instructors or TAs.

Question 1 (1 + 1 points): You move into a new neighborhood, and there are three other houses in your cul-de-sac. When talking to your new neighbors, lets call them Alice, Bob, and Cathy, you find out that Alice has three pets. Since you don't know if Alice is a cat or a dog lover, you can assume that the probability that Alice has a cat is $\frac{1}{2}$ and the probability that she has a dog is $\frac{1}{2}$. Alice reveals to you that one of her pets is a cat.

a) What is the probability that at least one of Alice's pets is a dog?

After getting to know your neighbors more, you find out the following: Alice has two dogs and a cat, Bob has a cat, a dog, and a hamster, while Cathy has four dogs. However, since you are busy with school, you never got a chance to see any of the pets. During your evening walk, you see a dog without a leash, and you want to alert your neighbors.

b) Whose door should you knock first and why?

Question 2 (1 + 1 + 1 + 2 points): Recall the 1-d Gaussian distribution that we studied in the class.

a) Prove that the Gaussian distribution is well normalized. In other words show that $\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = 1$

b) Formally show that for the Gaussian distribution

$$\mathbb{E}\left[x\right] = \mu$$
$$\mathrm{var}\left[x\right] = \sigma^2$$

c) Show that the maximum likelihood (ML) estimate for the Gaussian is given by

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n$$
$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} \left(x_n - \mu_{ML}\right)^2$$

d) Show that the ML estimate of the mean is unbiased but the variance is biased:

$$\mathbb{E}\left[\mu_{\mathrm{ML}}\right] = \mu$$
$$\mathbb{E}\left[\sigma^2_{\mathrm{ML}}\right] = \left(\frac{N-1}{N}\right)\sigma^2$$

Question 3 (1 + 1 + 1 points): a) Show that if X and Y are independent random variables, then

$$p\left(X,Y\right) = p\left(X\right)\cdot p\left(Y\right)$$
$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right]$$

b) Write some code to simulate a large number (say 10,000) of coin tosses with two independent coins, and empirically verify the above results. Write a short (5 - 10 sentence) summary of your experiment and what you observed.

c) Now let your coin tosses depend on each other (e.g., the value that coin 2 takes depends on the value of coin 1) and empirically verify that the above results do not hold. Write a short (5 - 10 sentence) summary of your experiment and what you observed.

Learning Outcomes After this homework you should

- Be able to gain a good understanding of Bayes rule and posterior probabilities.
- Become comfortable with the Gaussian distribution and its properties.
- Gain a deeper understanding of what independence of random variables mean.

Changelog

- 03 October 2016: Added learning outcomes
- 03 October 2016: Typo fix: missing comma in question 3.