Advanced Topics in Machine Learning 2015-2016

Yevgeny Seldin Christian Igel Brian Brost

Home Assignment 0

Deadline: Sunday, 6 August, 2015, 23:59

The assignments must be answered individually - each student must write and submit his/her own solution. We encourage you to work on the assignments on your own, but we do not prevent you from discussing the questions in small groups. If you do so, you are requested to list your group partners in your individual submission.

Submission format: *Please, upload your answers in a single* `.pdf` *file and additional* `.zip` *file with all the code that you used to solve the assignment. (The* `.pdf` *should* ***not*** *be part of the* `.zip` *file.)*

IMPORTANT: *We are interested in how you solve the problems, not in the final answers. Please, write down all your calculations.*

Question 1 (Probability theory refreshment)**.** An urn contains five red, three orange, and one blue ball. Two balls are randomly selected (without replacement).

1. What is the sample space of this experiment?
2. What is the probability of each point in the sample space?
3. Let X represent the number of orange balls selected. What are the possible values of X?
4. Calculate $\mathbb{P}\{X = 0\}$.
5. Calculate $\mathbb{E}[X]$.

Question 2 (Probability theory refreshment)**.** Let X and Y be two discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let p_X be the distribution of X, p_Y the distribution of Y, and p_{XY} the distribution of X and Y. In other words, $\mathbb{P}\{X = x\} = p_X(x)$, $\mathbb{P}\{Y = y\} = p_Y(y)$, and $\mathbb{P}\{(X = x) \text{ AND } (Y = y)\} = p_{XY}(x, y)$. A convenient way to represent a joint probability distribution of two discrete random variables is a table. For example, if X and Y are two fair coins then the joint distribution table looks like this:

$X \backslash Y$	0	1
0	1/4	1/4
1	1/4	1/4

. And if Z_1 and Z_2 are two fair coins and we define $X = Z_1 + Z_2$ and $Y = Z_1 \times Z_2$ then the joint distribution of X and Y is:

$X \backslash Y$	0	1
0	1/4	0
1	1/2	0
2	0	1/4

.

We remind you the following properties and definitions from the probability theory:

(a) $p_X(x) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$.

(b) If X and Y are independent then $p_{XY}(x, y) = p_X(x) p_Y(y)$.

(c) $\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p_X(x)$.

Starting from the definitions, prove the following identities:

1. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

2. If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. (Mark the step where you are using the independence assumption. Note that this assumption was not required in point 1.)

3. Provide an example of two random variables X and Y for which $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. (Describe how you define the random variables, provide a joint probability distribution table, and calculate $\mathbb{E}[XY]$ and $\mathbb{E}[X]\mathbb{E}[Y]$.)

4. $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$.

5. Variance of a random variable is defined as $\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$. Show that $\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$.

Question 3 (Markov's Inequality vs. Hoeffding's Inequality vs. Binomial Bound)**.** Let $X_1, \ldots, X_{10}$ be i.i.d. Bernoulli random variables with bias $\frac{1}{2}$. (I.e., $\mathbb{P}\{X_i = 0\} = \mathbb{P}\{X_i = 1\} = \frac{1}{2}$.)

1. Use Markov's inequality to bound the probability that $\sum_{i=1}^{10} X_i \geq 9$. (Hint: you have to define a new random variable $S = \sum_{i=1}^{10} X_i$ and apply Markov's inequality to S.)

2. Use Hoeffding's inequality to bound the probability of the same event. (Hint: now you do not need to group the variables together and you can exploit their independence.)

3. Calculate the exact probability of the above event. (Hint: S has Binomial distribution.)

4. Compare the three results.

Question 4 (Hoeffding's Inequality)**.**

1. An airline knows that any person making a reservation on a certain flight will not show up with a probability of 0.05 (5 percent). They introduce a policy to sell 100 tickets for a flight that can hold only 99 passengers. Bound the probability that the number of people that show up for a flight will be larger than the number of seats.

2. An airline has collected a sample of 10000 flight reservations and figured out that in this sample 5 percent of passengers who made a reservation on a certain flight did not show up. They introduce a policy to sell 100 tickets for a flight that can hold only 99 passengers. Bound the probability that the number of people that show up for a flight will be larger than the number of seats.

Question 5 (SVMs)**.** In Least Squares SVMS (LS-SVMs, Suykens Vandewalle, 1999) the hinge loss in the 2-norm soft margin SVM is replaced by the squared loss $L(y, y) = (y - y)^2$.

Please recall the Representer Theorem (e.g., see lecture notes Theorem 4.3 on page 39).

1. Is an approach via a Lagrangian really necessary?

2. What are the constraints on the objective variables?

Suykens, J.A.K.; Vandewalle, J. (1999) "Least squares support vector machine classifiers", Neural Processing Letters, 9 (3), 293-300.

Good luck!
Yevgeny, Christian, & *Brian*