Machine Learning 521289S
Spring 2019
Exercise 1: Basic Probability and Statistics & Bayesian Classification
Exercises 2 and 3a-d give exam points!

General Instructions

Use the upper case $P(\cdot)$ to denote the *probability mass function* (of a discrete random variable), and lower case $p(\cdot)$ to denote a *probability distribution* (of a continuous random variable). Vectors are written in bold lowercase letters ($\mathbf{x}$) and matrices in bold uppercase letters ($\mathbf{X}$). In handwriting, use bars over the letters to denote vectors ($\vec{x}$). Random variables are written in upper case letters and their values in lower case letters ($X = x$, $C = 1$).

One dimensional normal distribution with mean μ and variance σ^2 has the *probability density function* (PDF)

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Exercises

1. *Conditional Probability*

   Remember that the probability of an event A given the occurrence of event B is known as the *conditional probability* and is defined as

   $$\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)}, \tag{1}$$

   when $\Pr(B) > 0$. In case $\Pr(B) = 0$, the conditional probability is undefined. In the above, $\Pr(A|B)$ is the probability of A given B, $\Pr(A,B)$ is the probability that both events A and B occur (i.e. event $A \cap B$), and $\Pr(B)$ is the probability of event B. If the events are statistically independent, then $\Pr(A|B) = \Pr(A)$ and $\Pr(B|A) = Pr(B)$.

   (a) According to a statistical study in a population of 100 000 females, 89.835 % can expect to live to age 60, while 57.062 % can expect to live to age 80. Given that a woman is now 60 years old, what is the probability that she lives to age 80?

   (b) In a poker game, John has a very strong hand and bets 60 euros. The probability that Mary has a better hand is 0.04. If Mary had a better hand, she would raise with probability 0.9, but with poorer hand she would only raise with probability 0.1 (which includes bluffing and wrong judgement). Now, the question that is puzzling John is the following. If Mary raises, what is the probability that she has a better hand than John does?

2. *Bayes' Theorem*

   The Bayes decission rule uses the Bayes formula to make statistically ideal classifications[1].

[1] For more information, see Duda, Hart & Stork (2001) pages 20-25 (Chapter 2) or pages 1-3 in Chapter 2 of the old lecture notes (in Finnish).

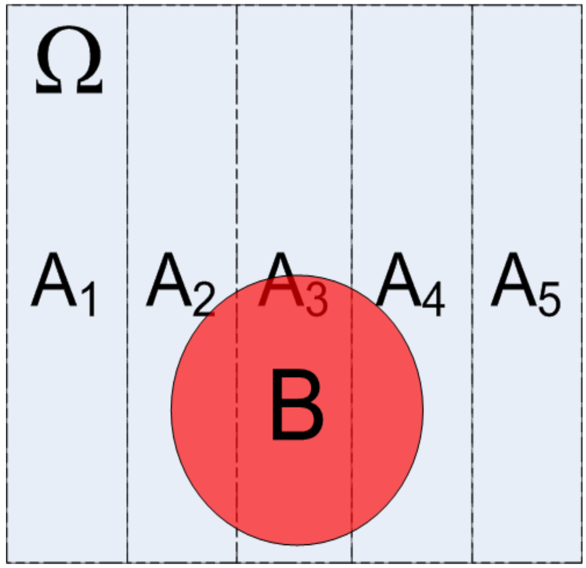

Figure 1: The space Ω is partitioned into disjoint events $A_1, \ldots, A_5$.

(a) Use the conditional probability formula to derive the *Bayes' Theorem*. In other words, you must show how the formula

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \tag{2}$$

can be obtained from the conditional probability formula.

(b) Let the probability space Ω be partitioned into disjoint subsets A_i, $i = 1, 2, \ldots, n$. In other words, let $\Omega = \cup_{i=1}^{n} A_i$, where $A_i \cap A_j = \emptyset$ whenever $i \neq j$. For example, see Figure 1 where $\Omega = A_1 \cup A_2 \cup A_5 \cup A_4 \cup A_5$. Now as the events A_i are mutually disjoint, the *Law of Total probability* states that

$$\Pr(B) = \Pr(B, A_1) + \Pr(B, A_2) + \Pr(B, A_3) + \Pr(B, A_4) + \Pr(B, A_5).$$

Generalize this idea and express the Bayes' Theorem (2) in the *alternative form*

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{k=1}^{n} \Pr(B|A_k)\Pr(A_k)}.$$

3. *Bayesian Classification Task*

Let us examine a simplified classification task where we have one feature $X \in \mathbb{R}$, a random variable describing e.g. the weight of a fish, and two classes $C \in \{1, 2\}$, e.g. two distinct fish species to be separated from each other. Let us further suppose that the both classes have equal prior probabilities ($P(C = 1) = P(C = 2) = \frac{1}{2}$), i.e. the fish species are equally abundant, and that for the class 1 the feature X is normally distributed with parameters $\mu_1 = 2$ and $\sigma_1^2 = 1$, and for class 2 it is normally distributed with $\mu_2 = 4$ and $\sigma_2^2 = 4$. In other words, $p(x|1) \sim N(2, 1^2)$ and $p(x|2) \sim N(4, 2^2)$.

(a) Sketch the class conditional probabilities $p(x|1)$ and $p(x|2)$ to the same coordinate frame as a function of x.

(b) Calculate the posterior probabilities $P(C=1|x)$ and $P(C=2|x)$ and sketch the them to the same coordinate frame as a function of x. Compare with the figure drawn at the part (a).

(c) Use the figure you have drawn in (b) to deduce the approximate thresholds for a sample being classified to the class 1 using the Bayes decision rule. What is the classification when the feature has value 3?

(d) Calculate the exact feature threshold value(s) to use in classification!

(e) What is the probability of making an error using this classifier?

4. *Marginal Distributions and Statistical Independence*

Table 1 shows data relating to a smoking habit and the prevalence of cancer collected from a group (Ω) of 60 individuals. Suppose that a person (ω) is chosen randomly from the group. Let us define a discrete random variable C so that $C(\omega)=1$ if this person has cancer and $C(\omega)=0$ if not. In addition, let S be another random variable describing the smoking habit, i.e. $S(\omega)=1$ if this person smokes and $S(\omega)=0$ if not. Together, these variables form a two-dimensional random variable (C,S) which has a joint *probability mass function* $P(c,s)=\Pr(\text{"}C=c \text{ and } S=s\text{"})$. Table 1 allows us to estimate estimate that the probability that the randomly selected individual does not smoke and does not have cancer is $P(0,0)=\frac{40}{60}=\frac{2}{3}$, and so forth.

	not smoke	**smoke**	total
no cancer	40	10	50
cancer	7	3	10
total	47	13	60

Table 1: Smoking and cancer.

Calculate the marginal distributions for both variables. In other words, marginalise $P(c,s)$ to find $P(c)$ that gives us the probability of having cancer when we do not know whether the person smokes or not, and $P(s)$ that tells us the odds of having a smoking habit.

Answers:
1. a) approx. 63%. b) approx. 27%. Hint: John makes the assessment here.
2. Hint: Law of total probability.
3. Hint: Sketch!