Course 67564	2022/23

Bayesian Machine Learning

Exercise 0: Linear Algebra and Probability

Prof. Yair Weiss — *TA: Roy Friedman*

Deadline: November 10, 2022

1 Questions

1.1 Linear Algebra

As you probably already noticed, we will use a lot of linear algebra in this course. Specifically, we will typically write everything in vector notation, since it will allow us to package many numbers together. Since this material isn't used in a lot of courses, it will be a good idea to practice these concepts before we use them.

1. Throughout the course, we will have many terms in the quadratic form $x^T Rx$, where x is a vector and R is some matrix. Find the derivative of the following function with respect to x:
$$f_1(x) = (x-\mu)^T R(x-\mu) \tag{1}$$
where $x, \mu \in \mathbb{R}^n$ and $R \in \mathbb{R}^{n\times n}$. Will the derivative have a different form if we know that R is symmetric?

2. Consider the following expression:
$$f_2(\theta) = \sum_{i=1}^{n} \left(h_i^T\theta - y_i\right)^2 \tag{2}$$
where $\theta \in \mathbb{R}^q$, $h_i \in \mathbb{R}^q$ and $y_i \in \mathbb{R}$. Show that this expression can be rewritten as:
$$f_2(\theta) = \|H\theta - y\|^2 \tag{3}$$
with some matrix $H \in \mathbb{R}^{n\times q}$ and some vector $y \in \mathbb{R}^n$

3. Consider the following function:
$$f_3(\theta, \lambda) = -C\log\frac{1}{\lambda} - \frac{1}{2}\lambda\sum_{i=1}^{n}\left(h_i^T\theta - y_i\right)^2 \tag{4}$$
where $y_i \in \mathbb{R}$, $h_i, \theta \in \mathbb{R}^q$ and $\lambda, C \in \mathbb{R}_+$. You can assume that this function is concave and has only one maximum with respect to θ and λ when $\lambda > 0$ (and it really is).
   (a) Find the θ that maximizes $f_3(\cdot)$. Does it depend on λ?
   (b) Find the λ that maximizes $f_3(\cdot)$. Does it depend on θ?[1]
   (c) Write down the pair θ and λ that maximize $f_3(\cdot)$
   (d) (Optional) To show that a function is convex, it is enough to show that the *Hessian* of the function at any point is a PSD matrix. As a reminder, the Hessian of a function from $\mathbb{R}^n$ to $\mathbb{R}$ is the Jacobian of the gradient of the function (a generalization of the second derivative). Show that $g(\theta, \lambda) = -f_3(\theta, \lambda)$ is a convex function for any $\lambda > 0$ and conclude that $f_3(\theta, \lambda)$ is concave for any $\lambda > 0$.

[1] Hint: use $\log\frac{1}{\lambda} = -\log\lambda$ to make differentiating easier

1.2 Probability

A probabilistic view is the core backbone of the course, so having a good intuition for stochastic events and random variables will be valuable.

4. Suppose we have some way to predict whether it will rain tomorrow using some fancy machine. We know that the probability that it will predict rain when there actually won't be any is p_{FP} (the "FP" stands for *false positive*). Conversely, the probability it predicts dry conditions when there will be rain is p_{FN} (the "FN" stands for *false negative*). We also somehow know that the probability that tomorrow will be rainy is p_r. What is the probability that there will be rain if our machine said there won't be?

5. Let x be a uniform random variable on the segment $\left[m-\frac{d}{2}, m+\frac{d}{2}\right]$ for some number m and $d>0$. The PDF of x is given by:
$$p(x)=\begin{cases}\frac{1}{C} & x\in\left[m-\frac{d}{2}, m+\frac{d}{2}\right]\\ 0 & \text{otherwise}\end{cases} \tag{5}$$
where C is some constant. We will denote this distribution as:
$$x\sim\mathcal{U}\left(\left[m-\frac{d}{2}, m+\frac{d}{2}\right]\right) \tag{6}$$
   (a) Find C
   (b) Find the mean and variance[2] of x
   (c) Let $y=x+\delta$ where δ is some (constant) number. What will be the distribution, mean and variance of y?

6. Let x and y be continuous random vectors that are independent (i.e. $p(x,y)=p(x)p(y)$). Show that:
$$\mathbb{E}\left[xy^T\right]-\mathbb{E}[x]\,\mathbb{E}\left[y^T\right]=0 \tag{7}$$
and conclude that:
$$\operatorname{cov}[x+y]=\operatorname{cov}[x]+\operatorname{cov}[y] \tag{8}$$
for any independent random vectors x and y

7. Let $x\in\mathbb{R}^n$ and $\eta\in\mathbb{R}^q$ be independent random vectors and let $H\in\mathbb{R}^{q\times n}$ be some constant matrix. Find the covariance of:
$$y=Hx+\eta \tag{9}$$
You can assume that $\mathbb{E}[x]$, $\mathbb{E}\left[xx^T\right]$ and $\operatorname{cov}[\eta]$ are known ahead of time[3]

2 Submission Guidelines

Submit a single PDF file which contains your answers to the questions in the exercise. In general, it is better if you type your homework, but if you prefer handwriting your answers, please make sure that the text is readable when you scan it.

Part of your assignment will be automatically graded in a Moodle quiz, at this link. These answers will be graded automatically, so write only numeric values where needed.

[2] Use the integral $\int_a^b x^n dx=\frac{1}{n+1}\left(b^{n+1}-a^{n+1}\right)$ for $n\neq 0$
[3] Hint: use the identity $\operatorname{cov}[y]=\mathbb{E}\left[yy^T\right]-\mathbb{E}[y]\,\mathbb{E}[y]^T$