MATH 462 ASSIGNMENT 4
VERSION November 14, 2023

ADAM M. OBERMAN

Instructions. You are also free to consult the textbook references or the internet, but be careful of conflicting notation and definitions. Submit your work on the MyCourses course page.

Math Solutions are preferred handwritten, but typed (good quality latex) is permitted. You can get help from other students, but you should do the write up yourself. Code solutions can be a PDF of a workbook, you can use any language.

4.1. **Analytic Geometry and Covariance matrices.**

Exercise 4.1 (Exercise 3.5 from MLL[1]). Consider the Euclidean vector space $\mathbb{R}^5$. A subspace $U \subseteq \mathbb{R}^5$ and $\boldsymbol{x} \in \mathbb{R}^5$ are given by

$$U = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 5 \\ 0 \\ 7 \end{bmatrix} \right\}, \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}$$

(a) Determine the orthogonal projection $\pi_U(\boldsymbol{x})$ of $\boldsymbol{x}$ onto U
(b) Determine the distance $d(\boldsymbol{x}, U)$

Exercise 4.2 (Inner products). (a) Rewrite the definition of an inner product on a vector space.
(b) Given the $n \times n$ matrix M, which is full rank, verify from the definition that $\langle x, y \rangle_M = (Mx)^\top (My)$ defines an inner product on $\mathbb{R}^n$. What goes wrong if the matrix has a non-trivial null-space?
(c) Give an example of a norm on $\mathbb{R}^n$ which does not come from an inner product.

Exercise 4.3 (Covariance Matrix). Let $n = 2$. Find the covariance matrix for the following datasets, S^m.

(a) $S^m = \{(1,1), (-1,-1), (1,0), (-1,0), (-1,1), (1,-1), (0,1), (0,-1)\}$
(b) $S^m = \{(t,t), (-t,-t), (1,0), (-1,0), (-1,1), (1,-1), (0,1), (0,-1)\}$, for any $t \in \mathbb{R}$.

Hint: recall the facts that for a vector $v = (a,b)$, $v^\top v = a^2 + b^2$ is a scalar, and $vv^\top = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$ is a matrix.

Exercise 4.4 (Covariance Matrix Theory). Prove that the covariance matrix, C, is symmetric and non-negative definite, meaning $x^\top C x \geq 0$ for all x. Assuming that the matrix is invertible, prove it is (strictly) positive definite.

4.2. **k-means Clustering.**

Exercise 4.5. For the following two dimensional data sets, plot the data by hand and indicate the clusterings by drawing a circle around the points in each cluster

(a) The dataset

$$\begin{aligned} S = \{&(.8,1), (1.2,1), (1,.8), (1,1.2), (-.8,-1), (-1.2,-1), (-1,-.8), (-1,-1.2), \dots \\ &(-.8,1), (-1.2,1), (-1,.8), (-1,1.2), (.8,-1), (1.2,-1), (1,-.8), (1,-1.2)\}. \end{aligned}$$

(b) $S = \{(3,1), (-3,-1), (-3,1), (3,-1)\}$. Indicate two possible clusterings.

Exercise 4.6. Consider the dataset $S = \{-5, -4, -3, 6, 8, 10\} \subset \mathbb{R}$.

(a) Starting from $W^0 = (w_1, w_2) = (6, 9)$ perform the k-means algorithm with $k = 2$ until a fixed point is reached.
(b) Let h_W be the minimizer of the k-means loss. Plot $h_W(x)$ on the interval $[-5, 10]$.

[1]Deisenroth, Marc Peter, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for machine learning. Cambridge University Press, 2020.

4.3. **Hypothesis classes for unsupervised ML.**

Exercise 4.7. Given a dataset $S = \{x_1, \ldots, x_m\}$ in $\mathbb{R}^d$. Given a hypothesis class $\mathcal{H}$ of functions $h : \mathbb{R}^d \to \mathbb{R}^d$, consider the loss, $L(h, S) = \frac{1}{m} \sum_{i=1}^{m} \|h(x_i) - x_i\|^2$.

(a) Let $\mathcal{H}_{lin} = \{h(x) = Mx \mid M \in \mathbb{R}^{d\times d}\}$ be the hypothesis class consisting of linear maps on $\mathbb{R}^d$. Assume S is mean zero, $\bar{x} = \frac{1}{m} \sum x_i = 0$. Find the minimizer, $\min_{h \in \mathcal{H}_{lin}} L(h, S)$. What is the corresponding M?

(b) Suppose $\tilde{S}$ is a dataset with nonzero mean, $\bar{x}$. Let $\mathcal{H}_{aff} = \{h(x) = Mx + b \mid M \in \mathbb{R}^{d\times d}, b \in \mathbb{R}^d\}$. Find the minimizer of $L(h, \tilde{S})$ over $\mathcal{H}_{aff}$. this question is a bit silly, the question should have been : $h(x) = b$, then $b = \mu$ and loss is the variance.

Exercise 4.8. Given a dataset $S = \{x_1, \ldots, x_m\}$ in $\mathbb{R}^d$, which is mean zero, $\bar{x} = \frac{1}{m} \sum x_i = 0$. Given a hypothesis class $\mathcal{H}$ of functions $h : \mathbb{R}^d \to \mathbb{R}^d$, consider the loss $L(h, S) = \frac{1}{m} \sum_{i=1}^{m} \|h(x_i) - x_i\|^2$. In this exercise we identify algorithms in terms of this loss and a specific hypothesis class.

(a) Let $\mathcal{H}_{lin,k} = \{h(x) = W^\top W x \mid W \in \mathbb{R}^{d\times k}\}$. Show that this class includes the projection matrices onto rank k subspaces. How do you describe the W in this case? (Hint: orthgonality). If we minimize the loss $L(h, S)$ over this hypothesis class, which familiar algorithm do we obtain?

(b) Let $\mathcal{H} = \{h(x) = \arg\min_{i=1}^{k} \|x - w_i\|^2 \mid w_i \in \mathbb{R}^d, i = 1, \ldots, k\}$. Describe the hypothesis class in words. Are the functions in $\mathcal{H}$ differentiable? Identify the corresponding algorithm.