Homework 6

Math 198: Math for Machine Learning

Due Date:
Name:
Student ID:

Instructions for Submission

Please include your name and student ID at the top of your homework submission. You may submit handwritten solutions or typed ones (LaTeX preferred). If you at any point write code to help you solve a problem, please include your code at the end of the homework assignment, and mark which code goes with which problem. Homework is due by start of lecture on the due date; it may be submitted in-person at lecture or by emailing a PDF to both facilitators.

1 Ridge Regression and Kernel Trick

1. (Adapted from CS189 Fa19 HW2.) Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be a data matrix, $\mathbf{y} \in \mathbb{R}^n$ be an observation vector, and $\mathbf{w}_\lambda \in \mathbb{R}^d$ be the ridge regression solution, i.e., $\mathbf{w}_\lambda = (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y}$. Furthermore, let $\mathbf{X} = \mathbf{U\Sigma V}^\top = \sum_{i=1}^{d} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ be the SVD of $\mathbf{X}$.

   (a) Show that $\mathbf{w}_\lambda = \sum_{i=1}^{d} \frac{\sigma_i}{\sigma_i^2+\lambda}\mathbf{v}_i\langle \mathbf{u}_i, \mathbf{y}\rangle$.

   (b) Deduce that the OLS solution $\mathbf{w}_{\text{OLS}} = \sum_{i=1}^{d} \frac{1}{\sigma_i}\mathbf{v}_i\langle \mathbf{u}_i, \mathbf{y}\rangle$.

   (c) Prove that $\lim_{\lambda\to 0} \mathbf{w}_\lambda = \mathbf{w}_{\text{OLS}}$.

   (d) Show that if $\mathbf{w}_\lambda \neq 0$, then the map $\lambda \to ||\mathbf{w}_\lambda||^2$ is strictly decreasing and strictly positive on $(0, \infty)$. What is the effect of λ on $\mathbf{w}_\lambda$?

2. Prove that the kernel trick holds for cubic polynomials in two variables. That is, if the feature map ϕ maps

$$\begin{bmatrix} a_i & b_i \end{bmatrix}^\top \mapsto \begin{bmatrix} a_i^3 & b_i^3 & \sqrt{3}a_i^2 b_i & \sqrt{3}a_i b_i^2 & \sqrt{3}a_i^2 & \sqrt{3}b_i^2 & \sqrt{6}a_i b_i & \sqrt{3}a_i & \sqrt{3}b_i & 1 \end{bmatrix}^\top$$

   then $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + 1)^3$.

2 Linear Algebra Review

1. Let V be an arbitrary vector space. Prove that the zero vector $\mathbf{0} \in V$ is unique. Additionally, prove that for any vector $\mathbf{v} \in V$, the additive inverse $-\mathbf{v}$ is unique.

2. Prove that the dot product is a valid inner product on $\mathbb{R}^n$.

3. Let V and W be arbitrary vector spaces. Prove that $\dim V = \dim W$ if and only if there exists an isomorphism $f : V \to W$.

4. Prove that trace is a linear map, i.e. $\text{tr}(c\mathbf{A} + \mathbf{B}) = c\text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.
5. Let $\mathbf{A}$ be a square matrix and λ an eigenvalue of $\mathbf{A}$. Prove that λ^k is an eigenvalue of $\mathbf{A}^k$.
6. (Adapted from CS189 Fa19 HW0.) Let $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$. Define $\mathbf{A} = \mathbf{v}\mathbf{w}^\top$. Find the non-zero eigenvalues of $\mathbf{A}$ and their eigenvectors, and determine the rank of the nullspace of $\mathbf{A}$.
7. Prove that a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is PSD if and only if there exists a matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ such that $\mathbf{A} = \mathbf{U}\mathbf{U}^\top$.