Introduction to Machine Learning

Fall 2024

University of Science and Technology of China

Lecturer: Jie Wang | Homework 7

Posted: December 24, 2024 | Due: January 2, 2024

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Singular Value Decomposition

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\operatorname{rank} \mathbf{A} = r$. The SVD of $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{m\times m}, \mathbf{\Sigma} \in \mathbb{R}^{m\times n}, \mathbf{V} \in \mathbb{R}^{n\times n}$, and we sort the diagonal entries of $\mathbf{\Sigma}$ in the descending order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. Denote

$$\mathbf{U_1} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r),\ \mathbf{U_2} = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m),$$
$$\mathbf{V_1} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r),\ \mathbf{V_2} = (\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n).$$

The column space of $\mathbf{A}$ is the set

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\}.$$

The null space of $\mathbf{A}$ is the set

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{A}\mathbf{y} = \mathbf{0}\}.$$

1. Show that
   (a) $P_{\mathcal{C}(\mathbf{A})}(\mathbf{x}) = \mathbf{U_1}\mathbf{U_1}^\top\mathbf{x}$;
   (b) $P_{\mathcal{N}(\mathbf{A})}(\mathbf{x}) = \mathbf{V_2}\mathbf{V_2}^\top\mathbf{x}$;
   (c) $P_{\mathcal{C}(\mathbf{A}^\top)}(\mathbf{x}) = \mathbf{V_1}\mathbf{V_1}^\top\mathbf{x}$;
   (d) $P_{\mathcal{N}(\mathbf{A}^\top)}(\mathbf{x}) = \mathbf{U_2}\mathbf{U_2}^\top\mathbf{x}$.
2. The Frobenius norm of $\mathbf{A}$ is

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}^2}.$$

   (a) Show that $\|\mathbf{A}\|_F^2 = \operatorname{tr}(\mathbf{A}^\top\mathbf{A})$.
   (b) Let $\mathbf{B} \in \mathbb{R}^{m\times n}$. Suppose that $\mathcal{C}(\mathbf{A}) \perp \mathcal{C}(\mathbf{B})$, that is,

$$\langle \mathbf{a}, \mathbf{b}\rangle = 0,\ \forall\, \mathbf{a} \in \mathcal{C}(\mathbf{A}),\ \mathbf{b} \in \mathcal{C}(\mathbf{B}).$$

   Show that

$$\|\mathbf{A} + \mathbf{B}\|_F^2 = \|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2.$$

Solution: ■

Exercise 2: Principle Component Analysis

Suppose that we have a set of data instances $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^d$. Let $\widetilde{\mathbf{X}} \in \mathbb{R}^{d\times n}$ be the matrix whose i^{th} column is $\mathbf{x}_i - \bar{\mathbf{x}}$, where $\bar{\mathbf{x}}$ is the sample mean, and $\mathbf{S}$ be the sample variance matrix.

1. For $\mathbf{G} \in \mathbb{R}^{d\times K}$, let us define

$$f(\mathbf{G}) = \mathrm{tr}(\mathbf{G}^\top \mathbf{S}\mathbf{G}). \tag{1}$$

   Show that $f(\mathbf{G}\mathbf{Q}) = f(\mathbf{G})$ for any orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{K\times K}$.

2. Please find $\mathbf{g}_1$ defined as follows by the Lagrange multiplier method.

$$\mathbf{g}_1 := \underset{\mathbf{g}\in\mathbb{R}^d}{\mathbf{argmax}}\{f(\mathbf{g}) : \|\mathbf{g}\|_2 = 1\}, \tag{2}$$

   where f is defined by (1). Notice that, the vector $\mathbf{g}_1$ is the first principal component vector of the data.

3. Please find $\mathbf{g}_2$ defined as follows by the Lagrange multiplier method.

$$\mathbf{g}_2 := \underset{\mathbf{g}\in\mathbb{R}^d}{\mathbf{argmax}}\{f(\mathbf{g}) : \|\mathbf{g}\|_2 = 1, \langle \mathbf{g}, \mathbf{g}_1\rangle = 0\},$$

   where $\mathbf{g}_1$ is given by (2). Similar to $\mathbf{g}_1$, the vector $\mathbf{g}_2$ is the second principal component vector of the data.

4. Please derive the first K principal component vectors by repeating the above process.

5. What is $f(\mathbf{g}_k)$, $k = 1, \ldots, K$? What about their meaning?

Solution:

■

Exercise 3: Properties of Transition Matrix

A transition matrix (also called a stochastic matrix, probability matrix) is a square matrix used to describe the transitions of a Markov chain. Each of its entries is a nonnegative real number representing a probability. A right (left) transition matrix is a square matrix with each row (column) summing to one. Without loss of generality, we study the right transition matrix in this exercise. Suppose that $\mathbf{T} \in \mathbb{R}^{n \times n}$ is a right transition matrix.

1. Show that 1 is an eigenvalue of $\mathbf{T}$.
2. Let λ be an eigenvalue of $\mathbf{T}$. Show that $|\lambda| \leq 1$.
3. Show that $\mathbf{I} - \gamma\mathbf{T}$ is invertible, where $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix and $\gamma \in (0, 1)$.

Solution:

■

Exercise 4: Planning with a Two-Armed Bandit

Consider a two-armed bandit with two states as shown in Figure 1. A player can either pull the Bandit 1 or Bandit 2 trigger, and the bandit will dispense coins and transit its state according to the following rules.

- **At State 1**, only Bandit 1 dispenses 1 coin. Pulling Bandit 1 does not cause a state transition, and pulling Bandit 2 has a $p_1 = 0.4$ probability of transitioning to State 2.
- **At State 2**, Bandit 1 dispenses 2 coins, and Bandit 2 dispenses 3 coins. Pulling Bandit 1 does not cause a state transition, and pulling Bandit 2 has a $p_2 = 0.8$ probability of transiting to State 1.

Now assume that the reward equals the number of coins dispensed, and the player can play the bandit infinite times.

1. Please find the state space $\mathcal{S}$, the action space $\mathcal{A}$, and the transition function $P(s'|s,a)$ of the two-armed bandit, and draw the Markov process diagram.
2. Let $\gamma = 0.9$. Please find the state value functions $V^\pi(s)$ for the given policy $\pi(a|s)$:
   (a) π_1: Always pull the Bandit 2.
   (b) π_2: Pull Bandit 2 at State 1, and pull Bandit 1 at State 2.
3. For the cases where $\gamma = 0.1$ and $\gamma = 0.99$, please find the optimal policy π^* and its state value function $V^{\pi^*}(s)$. Please explain the effect of the value of γ based on the results.

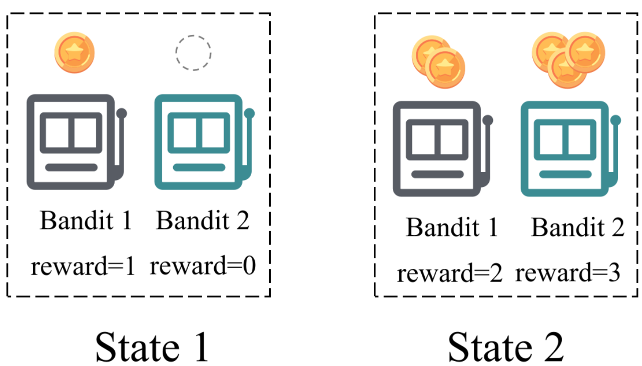

Figure 1: Illustration of the two armed-bandit.