CS4641 Machine Learning - Homework 1

Bo Dai

Deadline: 01/29 Wed, 23:59 PM

- Submit your answers as one single PDF file on Gradescope. **IMPORTANT: The solution to each problem/subproblem must be on a separate page. When submitting to Gradescope, please make sure to mark the page(s) corresponding to each problem/subproblem.**
- You will be allowed 2 total late days (48 hours) without penalty for the entire semester. Once those days are used, you will be penalized according to the following policy:
  - Homework is worth full credit before the due time.
  - It is worth 75% credit for the next 24 hours.
  - It is worth 50% credit for the second 24 hours.
  - It is worth zero credit after that.
- You are required to use Latex, or word processing software, to generate your solutions to the written questions. Handwritten solutions WILL NOT BE ACCEPTED.

1 Linear Algebra

Show that the matrix

$$P = \Phi(\Phi^T \Phi)^{-1} \Phi^T$$

acts as a **projection matrix**, meaning that for any vector $v \in \mathbb{R}^m$, it produces a new vector Pv that is the **projection** of v onto the space spanned by the columns of Φ. To verify this, demonstrate the following properties:

1. [**10 points**]

   Show that applying P twice is the same as applying it once (i.e., $P^2 = P$). This means that once a vector is projected, projecting it again does not change it. This property is called *Idempotence*.

2. [**10 points**]

   Show that P is symmetric, meaning $P^T = P$. This property is called *Symmetry*.

3. [**20 points**]

   For any vector v, show that $\Phi^T(v - Pv) = 0$. This means the difference between v and Pv is orthogonal to each column of Φ.

2 Probability and Statistics

1. **[15 points]**

   Let X and Y be random variables. The conditional expectation of X given Y, denoted as $E_X[X|Y]$, is a random variable that depends on Y:

   $$E_X[X|Y=y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx,$$

   where $f_{X|Y}(x|y)$ is the conditional density function of X given $Y = y$.

   In simpler terms, $E_X[X|Y=y]$ gives the expected value of X when Y takes a specific value y, while $E_X[X|Y]$ is a random variable that depends on the observed value of Y.

   Prove the following by the above definitions:

   (a) The law of total expectation: $E_Y[E_X[X|Y]] = E[X]$.

   (b) If X and Y are independent, then $E_X[X|Y] = E[X]$.

   (c) $E_Z[E_X[X|Y,Z]|Y] = E_X[X|Y]$.
   (**Hint: $E_x[X|Y,Z]$ is a random variable of both Y and Z.**)

2. **[15 points]**

   The Kullback-Leibler (KL) Divergence $D_{KL}(P\|Q)$ between two distributions $p(x)$ and $q(x)$ is defined as:

   $$D_{KL}(p\|q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

   What's the KL Divergence between two Gaussian distribution $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$?

3 Optimization

Suppose we want to minimize the function:

$$F(x, y) = y + (y - x)^2.$$

The actual minimum is $F = 0$ at $(x^*, y^*) = (0, 0)$. Solve the following questions in *vector* form.

1. **[15 points]**

   Find the gradient vector ∇F at the starting point $(x_0, y_0) = (1, 1)$.

2. **[15 points]**

   For full gradient descent (not stochastic) with step $s = \frac{1}{2}$, determine the updated point (x_1, y_1).