STAT 37710: Machine Learning	Autumn 2022

Homework 0 Solutions

Please do not distribute.

1 Probability and statistics

1. Bayes' theorem (*15 points*)

After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 100,000 people. What are the chances that you actually have the disease?

Solution:

Denote the event of having the disease as D and the event of testing positive as P. We have $\mathbb{P}(P|D) = 0.99$, $\mathbb{P}(P^c|D^c) = 0.99, \mathbb{P}(D) = 10^{-5}$. *We need to find* $\mathbb{P}(D|P)$.

$\mathbb{P}(D|P) = \frac{\mathbb{P}(D \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P|D)\mathbb{P}(D)+\mathbb{P}(P|D^c)\mathbb{P}(D^c)} = \frac{0.99\times10^{-5}}{0.99\times10^{-5}+0.01\times(1-10^{-5})} = ?$

2. Mean and variance (*15 points*) Let $X_1, X_2, \dots, X_n$ be i.i.d random variables distributed according to $\mathcal{N}(\mu, \sigma^2)$.

a.(*5 points*) Find two numbers a, b such that $aX_1 + b \sim \mathcal{N}(0, 1)$.

b.(*5 points*) Compute $\mathbb{E}[X_1 + 2X_2]$ and $\mathrm{Var}[X_1 + 2X_2]$.

c.(*5 points*) Setting $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$, compute the mean and variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$.

Solution:

(a) $\frac{X_1-\mu}{\sigma} \sim N(0,1)$, *so* $a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma}$

(b) $E(X_1 + 2X_2) = 3\mu$, $\mathrm{Var}\,(X_1 + 2X_2) = 5\sigma^2$

(c) Mean is 0, *variance is* σ^2

3. Empirical distribution (*25 points*)

For any function $g\colon \mathbb{R} \to \mathbb{R}$ and random variable X with PDF $f(x)$, recall that the expected value of $g(X)$ is defined as $\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(y)f(y)\mathrm{d}y$. For a boolean event A, define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Fix $F(x) = \mathbb{E}\mathbb{I}\{X \le x\}$. Let $X_1, \dots, X_n$ be independent and identically distributed random variables with CDF $F(x)$. Define $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i \le x\}$ to be the empirical CDF. Note, for every x, that $\widehat{F}_n(x)$ is an empirical estimate of $F(x)$.

a.(*5 points*) For any x, what is $\mathbb{E}\widehat{F}_n(x)$?

b.(*15 points*) Fix an x. Show that $\operatorname{Var}\widehat{F}_n(x) = \frac{F(x)(1-F(x))}{n}$.

c.(*5 points*) Using your answer to b, show that for all $x \in \mathbb{R}$, we have $\mathbb{E}\Big(\widehat{F}_n(x) - F(x)\Big)^2 \leq \frac{1}{4n}$.

Solution:

(a) $F(x)$

(b) $\operatorname{Var}\hat{F}_n(x) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var} I(X \leq x) = \frac{1}{n^2} nF(x)(1-F(x)) = \frac{F(x)(1-F(x))}{n}$ *since* $I(X \leq x)$ *follows the Binomial distribution* $Bin(1, F(x))$

(c) *LHS is variance of* $\hat{F}_n(x)$. *Apply the inequality* $(a+b)^2 \geq 4ab$ *to RHS of (b).*

2 Linear algebra

4. Linear systems (*10 points*) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^\top$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$.

a.(*5 points*) What is Ac?

b.(*5 points*) What is the solution to the linear system $Ax = b$?

Solution: $Ac = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$

$x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$

5. Rank (*20 points*) Let $v_1, \ldots, v_n$ be a set of non-zero vectors in $\mathbb{R}^d$. Let $V = [v_1 \ v_2 \ \ldots \ v_n]$ be the vectors concatenated.

a.(*5 points*) What is the minimum and maximum rank of $\sum_{i=1}^n v_i v_i^\top$?

b.(*5 points*) What is the minimum and maximum rank of V?

c.(*5 points*) Let $A \in \mathbb{R}^{D \times d}$ for $D > d$. What is the minimum and maximum rank of $\sum_{i=1}^n (Av_i)(Av_i)^\top$?

d.(*5 points*) What is the minimum and maximum rank of AV? What if V is rank d? **Solution:**

(a) *This matrix is just* VV^T, *and its rank is the rank of* V^T *(can be checked by SVD). Minimum rank is 1 (achieved when all* v_i*'s are equal). Maximum rank is* $\min(d, n)$, *since the matrix* $V^T \in \mathbb{R}^{n \times d}$.

(b) *Again, minimum rank is 1, maximum rank* $\min(n, d)$.

(c) *This matrix is* AVV^TA^T, *so its rank is equal to the rank of* AV. *Minimum rank is 0 since* A *can be 0. Note that rank*$(AV) \leq \min(\text{rank}(A), \text{rank}(V))$, *and* $\text{rank}(A) \leq d$ *and* $\text{rank}(V) \leq \min(n, d)$. *Hence* $\text{rank}(AV)$ *is at most* d. *It is achieved when, for example,* V *has rank* d *and* $A = \begin{pmatrix} I_d \\ 0 \end{pmatrix}$

(d) *See above discussion.*

6. Vectorized notation (*15 points***)** For possibly non-symmetric $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^\top \boldsymbol{A} x + y^\top \boldsymbol{B} x + c$. Define

$$\nabla_z f(x, y) = \begin{bmatrix} \partial f_{z_1}(x, y) & \partial f_{z_2}(x, y) & \dots & \partial f_{z_n}(x, y) \end{bmatrix}^\top .$$

a.**(***5 points***)** Explicitly write out the function $f(x, y)$ in terms of the components $A_{i,j}$ and $B_{i,j}$ using appropriate summations over the indices.

b.**(***5 points***)** What is $\nabla_x f(x, y)$ in terms of the summations over indices *and* vector notation?

c.**(***5 points***)** What is $\nabla_y f(x, y)$ in terms of the summations over indices *and* vector notation?

Solution:

(a) $f(x, y) = \sum_{i=1}^n \sum_{j=1}^n (x_i A_{ij} x_j + y_i B_{ij} x_j) + c$

(b) $\nabla_x f(x, y)[i] = \sum_{j=1}^n A_{ij} x_j + \sum_{j=1}^n A_{ji} x_j + \sum_{j=1}^n B_{ji} y_j$, *so* $\nabla_x f(x, y) = (A + A^\top)x + B^\top y$.

(c) $\nabla_y f(x, y)[i] = \sum_{j=1}^n B_{ij} x_j$ *so* $\nabla_y f(x, y) = Bx$