Home Assignment 1 Solutions

STAT 154/254: Modern Statistical Prediction & Machine Learning

Ajay Sharma — Fall 2024

Problem 1: Let $A \in \mathbb{R}^{3\times 3}$ be the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix}.$$

Calculate the following for A (and describe your steps):

(a) The eigenvalues.

(b) A unit eigenvector corresponding to each eigenvalue.

(c) $\det(A)$ and $\mathrm{tr}(A)$.

(d) The inverse A^{-1}.

(e) The Frobenius norm $\|A\|_F$ and the operator norm $\|A\|_{\mathrm{op}}$.

Solution. To find the eigenvectors and eigenvaulues, we first find the characteristic polynomial:

$$p_A(\lambda) = \det \begin{pmatrix} 1-\lambda & 2 & 3 \\ 3 & 1-\lambda & 2 \\ 2 & 3 & 1-\lambda \end{pmatrix} = -\lambda^3 + 3\lambda^2 + 15\lambda + 18.$$

The eigenvalues of A are exactly the roots of p_A, which we compute to be

$$6, \quad -\tfrac{3}{2} \pm \tfrac{\sqrt{3}}{2} i.$$

To find the eigenvector v for the eigenvalue λ, we need to find a unit-vector solution to the linear system of equations $Av = \lambda v$. For $\lambda_1 = 6$, we find that this is

$$v_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^\top.$$

for $\lambda_2 = -\frac{3}{2} + \frac{\sqrt{3}}{2} i$ we find

$$v_2 = \frac{1}{2\sqrt{3}}\left(-1 - \sqrt{3}i,\, -1 + \sqrt{3}i,\, 2\right)^\top,$$

and for $\lambda_3 = -\frac{3}{2} - \frac{\sqrt{3}}{2} i$ we find

$$v_2 = \frac{1}{2\sqrt{3}}\left(-1 + \sqrt{3}i,\, -1 - \sqrt{3}i,\, 2\right)^\top.$$

Next we recall that the determinant and the trace are just the product and the sum of the eigenvalues, respectively, hence

$$\det(A) = 6\Big(-\tfrac{3}{2} + \tfrac{\sqrt{3}}{2}i\Big)\Big(-\tfrac{3}{2} - \tfrac{\sqrt{3}}{2}i\Big) = 18, \quad \operatorname{tr}(A) = 6 + \Big(-\tfrac{3}{2} + \tfrac{\sqrt{3}}{2}i\Big) + \Big(-\tfrac{3}{2} - \tfrac{\sqrt{3}}{2}i\Big) = 3.$$

The inverse of A is given by $\operatorname{adj}(A)^\top / \det(A)$, where $\operatorname{adj}(A)$ is the *adjugate matrix*: the matrix whose i, j entry is $(-1)^{i+j}$ times the determinant of the 2×2 submatrix which arises by deleting the ith row and jth column from A. In particular, we have

$$\operatorname{adj}(A) = \begin{pmatrix} -5 & 1 & 7 \\ 7 & -5 & 1 \\ 1 & 7 & -5 \end{pmatrix}, \quad A^{-1} = \frac{1}{18}\begin{pmatrix} -5 & 7 & 1 \\ 1 & -5 & 7 \\ 7 & 1 & -5 \end{pmatrix}.$$

Finally, we compute the desired matrix norms. Recall that the Frobenius norm of A is the square root of the sum of the squared norms of the eigenvalues of A, hence

$$\|A\|_F = \sqrt{6^2 + \Big|\tfrac{3}{2} + \tfrac{\sqrt{3}}{2}i\Big|^2 + \Big|\tfrac{3}{2} - \tfrac{\sqrt{3}}{2}i\Big|^2} = \sqrt{42}.$$

Recall that the operator norm is the largest norm of an eigenvalue of A, hence

$$\|A\|_{\mathrm{op}} = \max\Big\{6,\ \big|\tfrac{3}{2} + \tfrac{\sqrt{3}}{2}i\big|,\ \big|\tfrac{3}{2} - \tfrac{\sqrt{3}}{2}i\big|\Big\} = 6.$$

Problem 2: Express the standard basis vector

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

as a linear combination of the column vectors

$$a_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Solution. We want a vector $x = (x_1, x_2, x_3)^\top$ such that

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} + x_3 \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

In matrix notation, this is exactly $e_1 = Ax$, where A is as in Q1. Since we already showed that A is invertible and we computed A^{-1}, we conclude

$$x = A^{-1}e_1 = \frac{1}{18}\begin{pmatrix} -5 \\ 1 \\ 7 \end{pmatrix}.$$

Problem 3: Let $A \in \mathbb{R}^{3\times 3}$ be the matrix

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{pmatrix}.$$

(a) Determine the rank of A.

(b) Find the null space (kernel) of A.

(c) Find the column space (image) of A.

(d) Compute the projection matrix onto $\ker(A)$.

(e) Compute the projection matrix onto $\mathrm{col}(A)$.

Solution. To begin, notice that the first two columns of A are linearly independent and that the third column is the sum of the first two. Since the rank is just the dimension of the columnspace, we see $\mathrm{rank}(A) = 2$.

To find the nullspace, we first note that the third column being the sum of the first two means that $(1, 1, -1)^\top$ is in the nullspace of A. But the rank–nullity theorem tells us that the nullspace of A must have dimension $3 - \mathrm{rank}(A) = 1$, hence

$$\mathrm{null}(A) = \Big\{ (\alpha, \alpha, -\alpha)^\top : \alpha \in \mathbb{R} \Big\}.$$

To find the column space, we can simply take the span of the first two columns, hence

$$\mathrm{col}(A) = \mathrm{span}\Big\{ (1,2,3)^\top,\ (1,1,1)^\top \Big\} = \Big\{ (\alpha+\beta,\ 2\alpha+\beta,\ 3\alpha+\beta)^\top : \alpha, \beta \in \mathbb{R} \Big\}.$$

Now we find the projection matrices. Projection onto $\mathrm{null}(A)$ is just projection onto the subspace spanned by $v = (1, 1, -1)^\top$, hence

$$P_{\mathrm{null}(A)} = \frac{vv^\top}{v^\top v} = \frac{1}{3}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

Projection onto $\mathrm{col}(A)$ is just projection onto the subspace spanned by the vectors $(1, 2, 3)^\top$ and $(1, 1, 1)^\top$. Hence, we can set

$$C = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$$

and get

$$P_{\mathrm{col}(A)} = C\big(C^\top C\big)^{-1} C^\top = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

Problem 4: Let $a, \mu \in \mathbb{R}$. Evaluate the Gaussian-type integral

$$\int_{-\infty}^{\infty} \exp\Bigl(-\tfrac{1}{2}\, ax^2 + \mu x\Bigr)\, dx,$$

and specify for which values of a it converges. (Hint: recall the PDF of $Z \sim N(\mu, \sigma^2)$.)

Solution. If $a \leq 0$ then the integrand diverges so the integral equals ∞. If $a > 0$ then we simply complete the square to get

$$\int_{-\infty}^{\infty} \exp\bigl(-\tfrac{1}{2}ax^2 + \mu x\bigr)\, dx = \int_{-\infty}^{\infty} \exp\Bigl(-\tfrac{1}{2}a\bigl(x - \tfrac{\mu}{a}\bigr)^2 + \tfrac{\mu^2}{2a}\Bigr)\, dx = \exp\Bigl(\tfrac{\mu^2}{2a}\Bigr)\sqrt{\frac{2\pi}{a}}.$$

Problem 5: Let $Z \sim N(0, I_d)$ and fix $a \in \mathbb{R}^d$. Compute

$$\mathbb{E}\bigl[e^{a^\top Z}\bigr] \;=\; \mathbb{E}\Bigl[e^{\sum_{i=1}^d a_i Z_i}\Bigr].$$

Then, for $X = AZ + \mu$ with $A \in \mathbb{R}^{d \times d}$, $\mu \in \mathbb{R}^d$, compute the moment-generating function

$$M(\mu, a) \;=\; \mathbb{E}\bigl[e^{a^\top X}\bigr].$$

Solution. First consider $Z \sim N(0, I_d)$. For $a \in \mathbb{R}^d$, we use the independence of the coordinates $Z_1, \ldots, Z_d$ of Z to compute:

$$\mathbb{E}\bigl[e^{a^\top Z}\bigr] = \mathbb{E}\Bigl[\exp\bigl(\sum_{i=1}^d a_i Z_i\bigr)\Bigr] = \prod_{i=1}^d \mathbb{E}\bigl[e^{a_i Z_i}\bigr].$$

Note that each factor can be computed with the help of Q1, since

$$\mathbb{E}\bigl[e^{a_i Z_i}\bigr] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\bigl(-\tfrac{1}{2}x^2 + a_i x\bigr)\, dx = \exp\bigl(\tfrac{a_i^2}{2}\bigr).$$

Thus, we have

$$\mathbb{E}\bigl[e^{a^\top Z}\bigr] = \prod_{i=1}^d \exp\bigl(\tfrac{a_i^2}{2}\bigr) = \exp\Bigl(\tfrac{\|a\|_2^2}{2}\Bigr).$$

Next we set $X = AZ + \mu$. By rearranging and applying the calculation for Z, we conclude:

$$\mathbb{E}\bigl[e^{a^\top X}\bigr] = \mathbb{E}\bigl[e^{a^\top (AZ + \mu)}\bigr] = \mathbb{E}\bigl[e^{a^\top AZ}\bigr]\, e^{a^\top \mu} = \exp\Bigl(\tfrac{\|A^\top a\|_2^2}{2}\Bigr)\, e^{a^\top \mu} = \exp\Bigl(\tfrac{a^\top \Sigma a}{2} + a^\top \mu\Bigr),$$

where $\Sigma = AA^\top$.

Problem 6: Using the result of Problem 5, find

$$\nabla_a M(\mu, a) \quad \text{and} \quad \nabla_\mu M(\mu, a),$$

and express each in compact matrix form.

Solution. From the form of $M(\mu, a)$ found in Q2 and the results of Q4 below, we can compute:

$$\nabla_a M(\mu, a) = (\Sigma a + \mu) \exp\Bigl(\tfrac{a^\top \Sigma a}{2} + a^\top \mu\Bigr), \quad \nabla_\mu M(\mu, a) = a \exp\Bigl(\tfrac{a^\top \Sigma a}{2} + a^\top \mu\Bigr).$$

Problem 7: For $u \in \mathbb{R}^d$, compute $\nabla_u f(u)$ for each of:

- $f(u) = a^\top u = \sum_{i=1}^d a_i u_i$.
- $f(u) = \|u\|_2^2 = \sum_{i=1}^d u_i^2$.
- $f(u) = u^\top A u$, where $A \in \mathbb{R}^{d \times d}$ is symmetric.

Solution. We compute these by writing out the definition of each function as a summation and taking the partial derivative in each coordinate:

- If $f(u) = \sum_{i=1}^d a_i u_i$, then $\partial f / \partial u_k = a_k$ for all $k = 1, \ldots, d$, so $\nabla_u f(u) = (a_1, \ldots, a_d) = a$.
- If $f(u) = \sum_{i=1}^d u_i^2$, then $\partial f / \partial u_k = 2u_k$ for all $k = 1, \ldots, d$, so $\nabla_u f(u) = (2u_1, \ldots, 2u_d) = 2u$.
- If $f(u) = \sum_{i=1}^d \sum_{j=1}^d u_i u_j A_{ij}$, then $\partial f / \partial u_k = 2 \sum_{i=1}^d u_i A_{ik}$ for all $k = 1, \ldots, d$, so $\nabla_u f(u) = 2Au$.

Problem 8: Let $X \in \mathbb{R}^{n \times d}$, $y \in \mathbb{R}^n$, and $\lambda > 0$. Define

$$f(u) \;=\; \|Xu - y\|_2^2 \;+\; \lambda \, \|u\|_2^2, \quad u \in \mathbb{R}^d.$$

Find the unique minimizer u^* of f in closed form in terms of X, y, λ.

Solution. By using the results of Q7 and the chain rule, we can compute

$$\nabla_u f(u) = 2X^\top (Xu - y) + 2\lambda u = 2\bigl(X^\top X + \lambda I\bigr) u - 2X^\top y, \quad \nabla_u^2 f(u) = X^\top X + \lambda I.$$

Since $\lambda_{\min}(X^\top X + \lambda I) \geq \lambda > 0$, we see that f is strictly convex, hence its unique stationary point must be a global minimizer. Then we see that $u \in \mathbb{R}^d$ satisfies $\nabla_u f(u) = 0$ if and only if

$$u = (X^\top X + \lambda I)^{-1} X^\top y,$$

and that the inverse is always well-defined. Therefore, we have

$$\arg\min_{u \in \mathbb{R}^d} \bigl(\|Xu - y\|_2^2 + \lambda \|u\|_2^2\bigr) = (X^\top X + \lambda I)^{-1} X^\top y.$$

Problem 9: Let $X \sim N(\mu, \Sigma)$ be a d-dimensional Gaussian with mean μ and invertible covariance Σ. For a fixed $a \in \mathbb{R}^d$, define the scalar $Y = a^\top X$. Compute:

$$\mathbb{E}[Y], \quad \mathrm{Var}(Y), \quad \mathbb{E}[\,X \mid Y = y], \quad \mathrm{Cov}(X \mid Y = y).$$

Solution. We can easily check that the vector $(X^\top, a^\top X)^\top$ has a multivariate Gaussian distribution, with

$$\begin{pmatrix} X \\ a^\top X \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} \mu \\ a^\top \mu \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma a \\ a^\top \Sigma & a^\top \Sigma a \end{pmatrix}\Big).$$

Therefore, by the Gaussian conditioning formula, we get

$$(X \mid Y = y) \sim \mathcal{N}\Big(\mu + \Sigma a \, \frac{y - a^\top \mu}{a^\top \Sigma a}, \ \Sigma - \Sigma a \, \frac{a^\top \Sigma}{a^\top \Sigma a}\Big).$$

In particular, we have

$$\mathbb{E}[X \mid Y = y] = \mu + \Sigma a \, \frac{y - a^\top \mu}{a^\top \Sigma a}, \quad \mathrm{Cov}(X \mid Y = y) = \Sigma - \Sigma a \, \frac{a^\top \Sigma}{a^\top \Sigma a}.$$

Problem 10: Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Fix $q \in (0,1)$ with $nq \in \mathbb{Z}$.

(a) Compute $\mathrm{Var}\big(\frac{1}{n}\sum_{i=1}^n X_i\big)$.

(b) Compute $f_n(p,q) = \mathbb{P}\big(\sum_{i=1}^n X_i = nq\big)$.

(c) Show $\lim_{n\to\infty} \frac{1}{n} \log f_n(p,q) = q \log \dfrac{q}{p} + (1-q)\log \dfrac{1-q}{1-p}$.

Solution. The calculations are straightforward applications of the independence and of known properties of Bernoulli random variables:

- Since $\mathrm{Var}(X_1) = p(1-p)$,

$$\mathrm{Var}\Big(\tfrac{1}{n}\sum_{i=1}^n X_i\Big) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{p(1-p)}{n}.$$

- Using the formula of the probability mass function for a Binomial random variable, we get

$$f_n(p,q) = \mathbb{P}\Big(\sum_{i=1}^n X_i = nq\Big) = \binom{n}{nq} p^{nq}(1-p)^{n-nq}.$$

- By the previous part, we have

$$\tfrac{1}{n}\log f_n(p,q) = \frac{\log(n!)}{n} + q\log p + (1-q)\log(1-p) - \frac{\log((nq)!)}{n} - \frac{\log((n(1-q))!)}{n}.$$

  Now recall Stirling's approximation, that we have

$$x! \sim \sqrt{2\pi x}\Big(\frac{x}{e}\Big)^x, \quad \log(x!) \sim \tfrac{1}{2}\log(2\pi x) + x\log x - x \quad (x \to \infty).$$

  Hence

$$\frac{\log(n!)}{n} \sim \log n - 1, \quad \frac{\log((nq)!)}{n} \sim q\log n + q\log q - q, \quad \frac{\log((n(1-q))!)}{n} \sim (1-q)\log n + (1-q)\log(1-q$$

  Therefore

$$\tfrac{1}{n}\log f_n(p,q) \longrightarrow q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p} \quad (n \to \infty).$$

Problem 11: Let $x_1, \ldots, x_n$ be i.i.d. $\mathrm{Exp}(\lambda)$ with density $p_\lambda(x) = \lambda e^{-\lambda x}$.

(a) Write the log-likelihood $\ell(\lambda) = \log \prod_{i=1}^n p_\lambda(x_i)$.

(b) Differentiate $\ell(\lambda)$, set to zero, and solve for the MLE $\widehat{\lambda}$.

(c) Compute $\mathbb{E}[\widehat{\lambda}^{-1}]$ and $\mathrm{Var}(\widehat{\lambda}^{-1})$.

Solution. We make the following calculations.

- First, we simply use the formula for p_λ to get

$$\ell(\lambda) = \log\Big(\prod_{i=1}^n p_\lambda(x_i)\Big) = \sum_{i=1}^n \log\big(\lambda e^{-\lambda x_i}\big) = n\log\lambda - \lambda\sum_{i=1}^n x_i.$$

- The derivative of ℓ is

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i,$$

  so $\ell'(\lambda) = 0$ is equivalent to $\lambda^{-1} = \frac{1}{n}\sum_i x_i$. This shows the MLE is

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}.$$

- The inverse of the MLE is just an average of IID random variables, so we can compute:

$$\mathbb{E}[\widehat{\lambda}^{-1}] = \mathbb{E}\Big[\tfrac{1}{n}\sum_{i=1}^n x_i\Big] = \tfrac{1}{n}\sum_{i=1}^n \mathbb{E}[x_i] = \lambda^{-1},$$

$$\mathrm{Var}(\widehat{\lambda}^{-1}) = \mathrm{Var}\Big(\tfrac{1}{n}\sum_{i=1}^n x_i\Big) = \tfrac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(x_i) = \tfrac{1}{n}\,\lambda^{-2}.$$

Problem 12: Let $x_1, \ldots, x_n$ be i.i.d. $N(\mu, 1)$. Derive the likelihood-ratio test for

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu = 1$$

at level α. Then construct a symmetric $1 - \alpha$ confidence interval for μ.

Solution. The likelihood ratio is the following statistic:

$$T(x_1, \ldots, x_n) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-1)^2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}} = \prod_{i=1}^n \exp\Big(-x_i + \tfrac{1}{2}\Big) = \exp\Big(-\sum_{i=1}^n x_i + \tfrac{n}{2}\Big).$$

The likelihood ratio test is the test that rejects for extreme values of T. Noticing that T is just a monotone transformation of the sample average, we can equivalently write the likelihood ratio test as the test which rejects for extreme values of the sample average $\frac{1}{n}\sum_{i=1}^n x_i$.

In order to choose the rejection threshold $t \in \mathbb{R}$ to maintain α significance for the null hypothesis $H_0 : \mu = 0$, we must solve

$$\alpha = \mathbb{P}_{H_0}\Big(\tfrac{1}{n}\sum_{i=1}^n x_i \geq t\Big) = \mathbb{P}\big(N(0,1) \geq t\sqrt{n}\big).$$

If we write Φ for the standard Gaussian CDF, then the unique solution is $t = \Phi^{-1}(\alpha)/\sqrt{n}$. Therefore, the likelihood ratio test for $H_0 : \mu = 0$ versus $H_1 : \mu = 1$ at significance α is the test which rejects when

$$\tfrac{1}{n}\sum_{i=1}^n x_i \;\geq\; \frac{\Phi^{-1}(\alpha)}{\sqrt{n}}.$$

To construct a symmetric confidence interval for μ at level α we need to choose $t \in \mathbb{R}$ such that

$$\alpha = \mathbb{P}\Big(\big|\tfrac{1}{n}\sum_{i=1}^n x_i - \mu\big| \geq t\Big).$$

That is, the coverage probability must be exactly α. Notice that

$$\tfrac{1}{n}\sum_{i=1}^n x_i - \mu = \tfrac{1}{n}\sum_{i=1}^n (x_i - \mu) \sim N\big(0, \tfrac{1}{n}\big),$$

and that this distribution does not depend on μ. Thus, we can choose $t \in \mathbb{R}$ such that

$$\alpha = \mathbb{P}\big(|N(0,1)| \geq t\sqrt{n}\big),$$

and we see that the unique solution is $t = \Phi^{-1}(\frac{\alpha}{2})/\sqrt{n}$. Therefore, a symmetric confidence interval for μ at significance α is given by

$$\Big(\sum_{i=1}^n x_i - \tfrac{1}{\sqrt{n}}\Phi^{-1}(\tfrac{\alpha}{2}),\ \sum_{i=1}^n x_i + \tfrac{1}{\sqrt{n}}\Phi^{-1}(\tfrac{\alpha}{2})\Big).$$