STAT 7670 — Multivariate Statistics

Midterm Examination (Practice)

Instructions

- Closed book, one page of notes
- Clearly justify all answers and show key steps.
- State assumptions (e.g., multivariate normality, full rank conditions).
- You may use standard results proved in class or homework, but cite them explicitly.

Problem 1 — Wishart, Independence, and Decomposition (20 points)

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}_p(\mu, \Sigma)$, with $\Sigma \succ 0$. Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad A = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^\top, \qquad S = \frac{1}{n-1}A.$$

(a) Distributional statements (8 pts)

1. State the distribution of A (name it and give its parameters).
2. State the distribution of $\bar{X}$.
3. State whether $\bar{X}$ and A are independent.

Hint: Use an orthogonal decomposition of the $n \times p$ data matrix into the projection onto the span of $\mathbf{1}_n$ and its orthogonal complement.

(b) Quadratic form (6 pts)

Under $H_0 : \mu = \mu_0$, show that

$$Q := n(\bar{X} - \mu_0)^\top \Sigma^{-1} (\bar{X} - \mu_0) \sim \chi^2_p.$$

Explain why replacing Σ with S changes the distribution.

(c) Whitening argument (6 pts)

Let $Y_i = \Sigma^{-1/2}(X_i - \mu)$.

1. Show that $Y_i \sim \mathcal{N}_p(0, I)$.
2. Express A in terms of the Y_i.
3. Use this to give intuition for why A has a Wishart distribution.

Problem 2 — Schur Complement and Conditional MVN (20 points)

Let

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{p_1+p_2}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right),$$

with Σ_{22} invertible.

(a) Conditional distribution (10 pts)

Derive the conditional mean and covariance of $X_1 \mid X_2 = x_2$.

Hint: Either complete the square in the joint density or use the block inverse formula.
The conditional covariance is the Schur complement

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

(b) Regression connection (6 pts)

Show that

$$\mathbb{E}[X_1 \mid X_2] = \alpha + BX_2$$

and identify α and B. Interpret B as a best linear predictor coefficient.

(c) Independence criterion (4 pts)

Give a necessary and sufficient condition (in terms of blocks of Σ) for X_1 and X_2 to be independent.

Problem 3 — Multivariate Linear Regression and Hypothesis Testing (20 points)

Consider

$$Y = XB + E, \qquad E_i \sim \mathcal{N}_p(0, \Sigma), \quad iid, \qquad \operatorname{rank}(X) = q,$$

where Y is $n \times p$, X is $n \times q$, and B is $q \times p$.

(a) Estimation and distribution (8 pts)

1. Derive $\widehat{B}$.
2. Give the mean and covariance of $\mathrm{vec}(\widehat{B})$.

Hint: Use

$$\widehat{B} = (X^\top X)^{-1} X^\top Y$$

and

$$\mathrm{vec}(MNA) = (A^\top \otimes M)\mathrm{vec}(N).$$

(b) Testing $H_0 : CB = 0$ (8 pts)

Let C be $r \times q$ with rank r. Define

$$E = Y^\top (I - P_X) Y, \qquad H = Y^\top (P_X - P_{X_0}) Y,$$

where X_0 is the design matrix under H_0.

1. Write expressions for Wilks' Lambda, Pillai's Trace, and Lawley–Hotelling Trace.
2. Explain how these depend on eigenvalues of $E^{-1}H$.

Hint: Consider the generalized eigenvalue problem

$$Hv = \lambda E v.$$

(c) Geometry (4 pts)

Interpret P_X and $I - P_X$ geometrically, and explain what "signal" and "noise" correspond to in H and E.

Problem 4 — MANOVA and Canonical Variates (20 points)

One-way MANOVA with g groups. Let W and B denote within- and between-group SSCP matrices.

(a) Canonical directions (10 pts)

Show that the directions a that maximize

$$\frac{a^\top B a}{a^\top W a}$$

are eigenvectors of $W^{-1}B$.

Hint: Use a Rayleigh quotient with constraint $a^\top W a = 1$.

(b) Wilks' Lambda (6 pts)

Let $\lambda_1, \dots, \lambda_s$ be the nonzero eigenvalues of $W^{-1}B$.

Show that

$$\Lambda = \prod_{j=1}^{s} \frac{1}{1+\lambda_j}.$$

Interpret what large vs. small λ_j imply about group separation.

(c) Test statistic comparison (4 pts)

Give one conceptual reason why Pillai's trace may be preferred over Wilks' Lambda in some settings.

Problem 5 — Hotelling's T^2 and Invariance (20 points)

Let $X_1, \dots, X_n \sim \mathcal{N}_p(\mu, \Sigma)$.

(a) Affine invariance (8 pts)

Define

$$T^2 = n(\bar{X} - \mu_0)^\top S^{-1}(\bar{X} - \mu_0).$$

Show that the test based on T^2 is invariant under full-rank affine transformations $Y = AX + b$, where A is invertible.

Hint: Track how $\bar{X}$ and S transform under A.

(b) Simultaneous inference (6 pts)

From the $1 - \alpha$ confidence ellipsoid for μ, derive a simultaneous confidence interval for $a^\top \mu$.
Explain how to obtain simultaneous intervals for each coordinate μ_j.

(c) Bonferroni comparison (6 pts)

Describe when Bonferroni intervals are close to Hotelling-based intervals. Your explanation should reference correlation and geometry.

Problem 6 — PCA, SVD, and Whitening (20 points)

Let centered data matrix X_c be $n \times p$. Define

$$S = \frac{1}{n-1} X_c^\top X_c.$$

(a) SVD connection (8 pts)

Let $X_c = UDV^\top$ be the SVD.

1. Show that the eigenvalues of S are $d_j^2/(n-1)$.
2. Show that the loading vectors are columns of V.

Hint: Compute $S = \frac{1}{n-1} VD^\top DV^\top$.

(b) Rank-k reconstruction (6 pts)

Let V_k contain the first k loading vectors. Show that the rank-k approximation is

$$\widehat{X}_{c,k} = X_c V_k V_k^\top.$$

Explain geometrically what $V_k V_k^\top$ represents.

(c) PCA vs. Whitening (6 pts)

Explain the difference between:

1. Rotating to principal axes (PCA), and
2. Whitening (sphering) the data.

If $\Sigma = V \Lambda V^\top$, give the whitening transformation explicitly.

End of Examination