Homework 4, Stat 541: Due Friday, Oct 16, 2009, 12 noon

Linear Algebra and Linear Models

Student Name: (replace this with your name)

October 6, 2009

Instructions: Edit this LaTex file with your solutions and generate a PDF file from it. E-mail the PDF to the usual class gmail address.

For some questions one needs to understand basis changes and associated coordinate transformations. To brush up (or to finally really understand what that is), you may want to check the solutions of Homework 2, Problem 11. See also Strang's text on linear algebra, the Appendix "Linear Transformations, Matrices, and Change of Basis."

The usual honor code applies (see the class webpage). In particular, it is strictly prohibited to consult previous years' solutions.

1. Interpretation of eigendecompositions: You know that for any real symmetric $p \times p$ matrix $\mathbf{S}$ there exists an orthonormal basis $(\mathbf{u}_j)_{j=1...p}$ of eigenvectors and associated eigenvalues $(\lambda_j)_{j=1...p}$ which can be assumed in descending order w.l.o.g.: $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. State what it means for $\mathbf{u}_j$ to be an eigenvector with eigenvalue λ_j, both formally (state the definition) and geometrically (how does an eigenvector move under $\mathbf{S}$?). Comment on what it means if $\lambda_j = 1$ or $\lambda_j = -1$ or $\lambda_j = 0$.

   Answer:

2. Under the assumption that the vectors $\mathbf{u}_j$ form an orthonormal system of eigenvectors of $\mathbf{S}$, form the matrix $\mathbf{U} = (\mathbf{u}_1, ..., \mathbf{u}_p)$ and form the diagonal matrix Λ that has the eigenvalues in the diagonal (zero off-diagona). Express what orthonormality of the vectors $\mathbf{u}_j$ means

in terms of $\mathbf{U}$ as a matrix equation. Then express the p eigenvalue conditions with a single matrix equation in terms of $\mathbf{S}$, $\mathbf{U}$ and Λ. (This is standard textbook material.)

Answer:

3. Again under the assumption that the vectors $\mathbf{u}_j$ form an orthonormal system of eigenvectors of $\mathbf{S}$, form the matrices $\mathbf{P}_j = \mathbf{u}_j \mathbf{u}_j^T$ and explain what type of linear map they describe (refer to HW2). Then express $\mathbf{S}$ in terms of $\mathbf{P}_1$,..., $\mathbf{P}_p$ and λ_1,..., λ_p. (Hint: $\mathbf{S}$ is just a linear combination of the maps $\mathbf{P}_j$.)

   Answer:

4. What is the name of the property of $\mathbf{S}$ called that is equivalent to $\lambda_p \geq 0$? Same for $\lambda_p > 0$? (This is just regurgitating textbook material.)

   Answer:

5. What is the matrix of the linear transformation given by $\mathbf{S}$ in the new coordinate system after changing the basis to $\mathbf{u}_1$,..., $\mathbf{u}_p$?

   Answer:

6. The following is called the Rayleigh quotient of $\mathbf{S}$:

$$R(\mathbf{u}) \;=\; \frac{\mathbf{u}^T \mathbf{S} \mathbf{u}}{\|\mathbf{u}\|^2}$$

   Find and interpret the stationary equations for $R(\mathbf{u})$: $\partial_{\mathbf{u}} R(\mathbf{u}) = \mathbf{0}$ (∂_u = gradient w.r.t. $\mathbf{u}$).

   Answer:

7. Assume $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal eigenvectors of $\mathbf{S}$ with identical eigenvalues $\lambda_1 = \lambda_2$. Consider the following one-parameter family of linear combinations $\mathbf{u}(t) = \cos(t)\mathbf{u}_1 + \sin(t)\mathbf{u}_2$. What geometric shape is traced by $\mathbf{u}(t)$? What can you say about $\mathbf{u}(t)$ in relation to $\mathbf{S}$? In what relation is $\mathbf{u}(t)$ relative to the eigenvectors $\mathbf{u}_3, ..., \mathbf{u}_p$?

   Answer:

8. What is the eigendecomposition of an orthogonal projection of rank r? How unique is it?

   Answer:

9. What is the matrix of on orthogonal projection after a change to a basis of eigenvectors? (This question is almost redundant with the previous problem.)

   Answer:

10. A fact about traces: show that if $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times m$, then $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$.

    Answer:

11. Use the previous fact to show that traces do not change under basis/coordinate transformation.

    Answer:

12. Using the previous fact, what is the trace of $\mathbf{S}$ in terms of the eigenvalues? Why?

    Answer:

13. Using the previous fact, what is the trace of an orthogonal projection?

    Answer:

14. Properties of the diagonal elements of orthogonal projection matrices: Prove $0 \leq P_{ii} \leq 1$. To this end, write P_{ii} in terms of the matrix $\mathbf{U}$ and the eigenvalues λ_j according to Problem 2 above. Use what you know about the eigenvalues of projections.

    Answer:

15. Consider a simple linear regression where we subtracted the mean from the response vector and from the predictor vector.

    (a) Write down the projection matrix $\mathbf{P}$ onto the centered predictor vector $\mathbf{x} \in \mathrm{I\!R}^N$.

    (b) Comment on the "total budget" in the diagonal of $\mathbf{P}$.

    (c) Interpret the diagonal elements in terms of self-influence.

    (d) Interpret the diagonal elements in terms of variance of fitted values and residuals.

    (e) Interpret the off-diagonal elements in terms of correlation between pairs of fitted values and pairs of residuals.

    Answer:

16. If $\mathrm{V}[X]$ is the covariance matrix of a p-dimensional random vector X, express $\mathrm{Var}[\mathbf{a}^T X]$ in terms of $\mathrm{V}[X]$ and $\mathbf{a}$, and assume that $\mathbf{a}$ is a unit vector. If you know $\mathrm{V}[X]$, what do you know about the distribution of X in all directions?

    Answer:

17. Two non-negative matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are said to be ordered, "$\mathbf{S}_1 \geq \mathbf{S}_2$" if $\mathbf{S}_1 - \mathbf{S}_2$ is non-negative definite. If we are given two p-dimensional random vectors X and Y, and if $\mathrm{V}[X] \geq \mathrm{V}[Y]$, what does this mean in view of the previous problem?

    Answer:

18. In a linear regression problem let $\mathbf{x}_1$,...,$\mathbf{x}_N$ be the "feature vectors", that is, the rows of the predictor matrix written as $(p+1)$-dimensional column vectors. Assume that they are N i.i.d. samples from the $(p+1)$-dimensional random vector X (and you can even allow the first coordinate to be all ones for the intercept, thought of as i.i.d. samples from the pointmass at 1.0). Q: Is it desirable that $\mathrm{V}[X]$ be "large" or "small" in the sense of the ordering of symmetric matrices defined in the previous problems? If it helps use the notation $\mathbf{X} = (\mathbf{x}_1, ...\mathbf{x}_N)^T$ for the $N \times (p+1)$ predictor matrix. (No need to prove anything about ordering of inverse matrices; you can argue that point intuitively.)

    Answer:

19. What does adjustment of a $N \times p$ data matrix $\mathbf{X}$ with regard to $(1, ..., 1)^T \in \mathbb{R}^N$ mean?

    Answer:

20. Write down the projection matrix for a categorical variable with three categories and category sizes $N_1 + N_2 + N_3 = N$. Assume that the cases are listed category $k = 1$ first, then $k = 2$, then $k = 3$, so that you can use a block-diagonal notation with blocks of size $N_k \times N_k$. Finally, indicate what the matrix is that adjusts for this categorical variable.

    Answer: