Final Exam

You may not use any books, notes, or computer programs (e.g., Julia). Throughout this exam (with one exception noted below) we use standard mathematical notation; and in particular, we do not use (and you may not use) notation from any computer language, or from any strange or non-standard mathematical dialect (e.g., physics). (One problem, however, specifically involves the Julia language.)

Some questions are True/False. For these problems simply circle the appropriate response. You do not need to give any justification for your answers to these questions. We will give partial credit for multiple choice problems left with no answer. If we can’t tell which response you are selecting, we will give zero credit.

For the other problems you are asked for a free-form answer, which must be written in the space below the problem statement. You should do your scratch work on scratch paper.

All problems have equal weight. Some are easy. Others, not so much. We recommend doing the easy ones first.

Name: 

SUID: 

(For EE103 staff only)

<table>
<thead>
<tr>
<th>Problem</th>
<th>Score</th>
<th>Problem</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/10</td>
<td>7</td>
<td>/10</td>
</tr>
<tr>
<td>2</td>
<td>/10</td>
<td>8</td>
<td>/10</td>
</tr>
<tr>
<td>3</td>
<td>/10</td>
<td>9</td>
<td>/10</td>
</tr>
<tr>
<td>4</td>
<td>/10</td>
<td>10</td>
<td>/10</td>
</tr>
<tr>
<td>5</td>
<td>/10</td>
<td>11</td>
<td>/10</td>
</tr>
<tr>
<td>6</td>
<td>/10</td>
<td>12</td>
<td>/10</td>
</tr>
</tbody>
</table>

Total /120
1. *Matrix notation.* Suppose the block matrix

\[
\begin{bmatrix}
A & I \\
I & C
\end{bmatrix}
\]

makes sense, where \(A\) is a \(p \times q\) matrix. What are the dimensions of \(C\)?
2. Oh no! A question on convolution. What is $(1, -1) \ast (1, 0, 1)$?
3. Suppose $F^T G = 0$, where $F$ and $G$ are $n \times k$ matrices. Determine whether each of the following statements must always be true, or can be false. ‘Must be true’ means the statement holds for any $n \times k$ matrices $F$ and $G$ that satisfy $F^T G = 0$, without any further assumptions; ‘can be false’ means that there are $n \times k$ matrices $F$ and $G$ that satisfy $F^T G = 0$, but the statement does not hold. Circle the correct response. Partial credit will be given if you don’t circle either response.

(a) Either $F = 0$ or $G = 0$.  
True. False.

(b) The columns of $F$ are orthonormal.  
True. False.

(c) Each column of $F$ is orthogonal to each column of $G$.  
True. False.

(d) The matrices $F$ and $G$ are square or tall, i.e., $n \geq k$.  
True. False.

(e) The columns of $F$ are linearly dependent.  
True. False.
4. Averages and affine functions. Suppose that $G : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function. Let $x_1, \ldots, x_k$ be $n$-vectors, and define the $m$-vectors $y_1 = G(x_1), \ldots, y_k = G(x_k)$. Let

$$
\overline{x} = (x_1 + \cdots + x_k)/k, \quad \overline{y} = (y_1 + \cdots + y_k)/k
$$

be the averages of these two lists of vectors. (Here $\overline{x}$ is an $n$-vector and $\overline{y}$ is an $m$-vector.)

Is it true that we always have $\overline{y} = G(\overline{x})$? In words: Is the average of an affine function applied to a list of vectors always the same as the affine function applied to the average of the list of vectors? If so, give a short explanation. If not, give a simple counterexample. (That is, a specific list of $n$-vectors and a specific affine function $G$ for which $\overline{y} \neq G(\overline{x})$.) You can use the fact that an affine function $G : \mathbb{R}^n \to \mathbb{R}^m$ has the form $G(x) = Ax + b$, where $A$ is an $m \times n$ matrix and $b$ is an $m$-vector.
5. **Matrix with acute columns.** Suppose $A$ is an $m \times n$ matrix with nonzero columns $a_1, \ldots, a_n$, and in addition, any pair of columns makes an acute angle, i.e., $|\angle(a_i, a_j)| < 90^\circ$ for all $i, j = 1, \ldots, n$. What can you say about the entries of the Gram matrix $G = A^T A$?
6. **Constrained least squares in Julia.** You are asked to write some Julia code to compute the $\hat{x}$ that minimizes $\|Ax - b\|^2$ subject to $Cx = d$, where $A$ is an $m \times n$ matrix, $b$ is an $m$-vector, $C$ is a $p \times n$ matrix, and $d$ is a $p$-vector. These are given as the Julia quantities $A$, $b$, $C$, and $d$, and the dimensions $m$, $n$, and $p$ are given as $m$, $n$, and $p$. You are to put the value of $\hat{x}$ in $\text{xhat}$. (You can assume that the associated KKT matrix is invertible.)

Write two lines of Julia code below that carries this out. Your code should be simple and clear. You do not need to justify your answer.

**Hint.** Recall that the optimality conditions for this constrained least squares problem are

$$2A^T Ax + C^T z = 2A^T b, \quad Cx = d,$$

where $z$ is the vector of Lagrange multipliers.
7. Modifying a classifier. A co-worker develops a classifier of the form \( \hat{y} = \text{sign}(x^T \beta + v) \), with \( v < 0 \), where the \( n \)-vector \( x \) is the feature vector, and the \( n \)-vector \( \beta \) and scalar \( v \) are the classifier parameters. The classifier is evaluated on a given test data set. The false positive rate is the fraction of the test data points with \( y = -1 \) for which \( \hat{y} = 1 \). (We will assume there is at least one data point with \( y = -1 \).)

Are each of the following statements true or false? True means it always holds, with no other assumptions on the data set or model; false means that it need not hold. Circle the correct response below each statement. You do not need to justify your answer. Some credit will be given if you decline to answer each question. You can indicate this by writing ‘My attorney has recommended that I not respond to this question at this time’.

(a) Replacing \( v \) with zero will reduce, or not increase, the false positive rate.
   True. False.

(b) Replacing \( \beta \) with zero will reduce, or not increase, the false positive rate.
   True. False.

(c) Halving \( v \) (i.e., replacing \( v \) with \( v/2 \)) will reduce, or at least not increase, the false positive rate.
   True. False.

(d) Halving \( \beta \) (i.e., replacing \( \beta \) with \( (1/2)\beta \)) will reduce, or at least not increase, the false positive rate.
   True. False.
8. **Linear dynamical system with 2× down-sampling.** We consider a linear dynamical system with \( n \)-vector state \( x_t \), \( m \)-vector input \( u_t \), and dynamics given by

\[
x_{t+1} = Ax_t + Bu_t, \quad t = 1, 2, \ldots,
\]

where the \( n \times n \) matrix \( A \) and \( n \times m \) matrix \( B \) are given. Define \( z_t = x_{2t-1} \) for \( t = 1, 2, \ldots \), i.e.,

\[
z_1 = x_1, \quad z_2 = x_3, \quad z_3 = x_5, \ldots
\]

(The sequence \( z_t \) is the original state sequence \( x_t \) 'down-sampled' by \( 2 \times \).) Define the \((2m)\)-vectors \( w_t \) as \( w_t = (u_{2t-1}, u_{2t}) \) for \( t = 1, 2, \ldots \), i.e.,

\[
w_1 = (u_1, u_2), \quad w_2 = (u_3, u_4), \quad w_3 = (u_5, u_6), \ldots
\]

(Each entry of the sequence \( w_t \) is a stack of two consecutive original inputs.)

Show that \( z_t, w_t \) satisfy the linear dynamics equation \( z_{t+1} = Fz_t + Gw_t \), for \( t = 1, 2, \ldots \). Give the matrices \( F \) and \( G \) in terms of \( A \) and \( B \). Be sure to specify the dimensions of \( F \) and \( G \).
9. **Minimum force to move a mass to a given position.** A 1kg mass is initially at rest on a level frictionless surface. A force $f_1$ is applied for a one second period, then a force of $f_2$ for the second one second period, and so on, up to a force $f_{10}$ applied for the 10th one second period. We let the scalar $p$ denote the position of the mass after 10 seconds, and the 10-vector $f$ denote the vector of the applied force sequence. These are related by $p = a^T f$, where $a = (19/2, 17/2, \ldots, 1/2)$. Find the force vector $\hat{f}$ that satisfies $p = 1$ (i.e., moves the mass to the position $p = 1$) and minimizes $f_1^2 + \cdots + f_{10}^2$. Note that we are not constraining the velocity of the mass, hinting that this particular application is not socially positive. You do not have to give the solution numerically; you can express it in terms of the vector $a$ defined above, using standard vector and matrix notation.
10. *Nonlinear auto-regressive model.* We have a (scalar) time series $z_1, z_2, \ldots, z_T$. The following one step ahead prediction model is proposed:

$$\hat{z}_{t+1} = \theta_1 z_t + \theta_2 z_{t-1} + \theta_3 z_t z_{t-1},$$

where $\theta = (\theta_1, \theta_2, \theta_3)$ is the model parameter vector. The sum of the squares of the prediction error of this model on the given time series is

$$\sum_{t=2}^{T-1} (\hat{z}_{t+1} - z_{t+1})^2.$$  

(Note that we must start this sum with $t = 2$, since $z_0$ and $z_{-1}$ are not defined.) Express this quantity as $\|A\theta - b\|^2$, where $A$ is a $(T-2) \times 3$ matrix and $b$ is a $(T-2)$-vector. (You must say what the entries of $A$ and $b$ are. They can involve the known data $z_1, \ldots, z_T$.)

*Remark.* Finding $A$ and $b$ is the first step in fitting the parameters $\theta$ to the data. We are not asking you to find $\theta$, but only to set up the least squares problem you’d solve to carry out the least squares fit.
11. *Fitting models to two different but similar populations.* We wish to fit two different models to data from two different but similar populations, for example males and females. The models are given by \( \hat{y} = x^T \beta \) for the first group and \( \hat{y} = x^T \tilde{\beta} \) for the second group, where \( x \) is the \( n \)-vector of features, \( \hat{y} \) is the prediction, \( \beta \) is the \( n \)-vector of model parameters for the first group and \( \tilde{\beta} \) is the \( n \)-vector of model parameters for the second group. (We can include an offset in the two models by including a feature that is always one.)

Our training data consists of \( x^{(1)}, \ldots, x^{(N)} \) and \( y^{(1)}, \ldots, y^{(N)} \) from the first population, and \( \tilde{x}^{(1)}, \ldots, \tilde{x}^{(N)} \) and \( \tilde{y}^{(1)}, \ldots, \tilde{y}^{(N)} \) from the second group. (For simplicity we assume that we have an equal number of training data points in the two groups.)

Our main goal in choosing the parameter \( n \)-vectors \( \beta \) and \( \tilde{\beta} \) is that the sum of squares of the prediction errors for the first group (using the first model) and the sum of squares of the prediction errors for the second group (using the second model) is small. Our secondary objective is that the two parameter vectors \( \beta \) and \( \tilde{\beta} \) are not too different. (This desire is based on our idea that the two groups are similar, so the associated models should not be too different.)

Capture the goals expressed above as a bi-objective least squares problem with variable \( \theta = (\beta, \tilde{\beta}) \). Identify the primary objective \( J_1 = \|A_1 \theta - b_1\|^2 \) and the secondary objective \( J_2 = \|A_2 \theta - b_2\|^2 \). Give \( A_1, A_2, b_1 \) and \( b_2 \) explicitly. Your solution can involve the \( n \times N \) data matrices \( X \) and \( \tilde{X} \), whose columns are \( x^{(i)} \) and \( \tilde{x}^{(i)} \), respectively, and the two \( N \)-vectors \( y^d \) and \( \tilde{y}^d \), whose entries are \( y^{(i)} \) and \( \tilde{y}^{(i)} \), respectively.
12. Lecture attendance, laptops in lecture, and final grade. A study collects data on a large number of students in a lecture course, with the goal of predicting the effect (or at least the association) of lecture attendance and laptop use on the final exam grade. The regressor is the 2-vector $x$, where $x_1$ is the student's lecture attendance, expressed as a number between 0 and 1 (with, say, 0.78 meaning the student attended 78% of the lectures), and $x_2$ is a Boolean feature that codes whether or not the student routinely used a laptop during lecture, with $x_2 = 0$ meaning she did not, and $x_2 = 1$ meaning that she did. The outcome variable $y$ is the student's final exam grade, scaled to be between 0 and 100 points. A basic regression model $\hat{y} = x^T \beta + v$ is fit to the data, and checked with out-of-sample validation.

(a) Give a one sentence interpretation of $\beta_1$.

(b) Give a one sentence interpretation of $\beta_2$.

(c) Suggest a value of $\beta_1$ that would not surprise you, and give a one sentence explanation that might be appropriate if the value of $\beta_1$ were your value.

(d) Suggest a value of $\beta_2$ that would not surprise you, and give a one sentence explanation that might be appropriate if the value of $\beta_2$ were your value.

For parts (c) and (d) we are looking for a plausible guess of the values, along with a plausible story that would go along with your guessed value.