You have 80 minutes to complete this exam.
You may use notes or printouts from the course website,
but no electronic resources.

Circle your final answer for each question.
(1) What is the distance between the origin and the hyperplane $3x_1 + 2x_2 - x_3 = 0$?

(2) True or False. The union of any two convex sets is also convex. (Note: the union of two sets is the set of all points contained in either set.)

(3) A genetic counselor informs a patient that their odds of getting disease are 1 in 300. What is the probability the patient will get the disease?

(4) True or False. Every square matrix has a pseudoinverse.

(5) True or False. If your dataset contains either categorical inputs or a categorical response variable, you should use logistic regression rather than linear regression.
Part II (20 points)

You are given a set of inputs \((x)\) and responses \((y)\)

\[
\begin{array}{c|c}
 x & y \\
 0 & 2.1 \\
 1 & 3.1 \\
 2 & 5.8 \\
 3 & 11.3 \\
\end{array}
\]

You hypothesize that these data fit one of two models: \(y = \beta_0 + \beta_1 x\) or \(y = \beta_0 e^x\). You want to fit the parameters using linear regression.

Construct a design matrix for each model using the data in the above table. (The design matrix is the matrix \(X\) in the linear system \(y = X\beta + \epsilon\).)
You use `fitlm` to fit both models, with the following outputs.

**Linear regression model:**

\[ y \sim 1 + x + x^2 \]

**Estimated Coefficients:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.155</td>
<td>0.23974</td>
<td>8.9889</td>
<td>0.070533</td>
</tr>
<tr>
<td>x</td>
<td>-0.345</td>
<td>0.385</td>
<td>-0.8961</td>
<td>0.53485</td>
</tr>
<tr>
<td>x^2</td>
<td>1.125</td>
<td>0.12298</td>
<td>9.1476</td>
<td>0.069319</td>
</tr>
</tbody>
</table>

Number of observations: 4, Error degrees of freedom: 1
Root Mean Squared Error: 0.246
R-squared: 0.999, Adjusted R-Squared 0.996
F-statistic vs. constant model: 421, p-value = 0.0344

**Linear regression model:**

\[ y \sim 1 + \exp x \]

**Estimated Coefficients:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.8553</td>
<td>0.25642</td>
<td>7.2353</td>
<td>0.018572</td>
</tr>
<tr>
<td>exp x</td>
<td>0.47699</td>
<td>0.023746</td>
<td>20.087</td>
<td>0.0024692</td>
</tr>
</tbody>
</table>

Number of observations: 4, Error degrees of freedom: 2
Root Mean Squared Error: 0.355
R-squared: 0.995, Adjusted R-Squared 0.993
F-statistic vs. constant model: 403, p-value = 0.00247

Which model would give better predictions? Why?
Part III (30 points)

Convert the following linear program to standard form:

\[
\begin{align*}
\text{max } f &= 2x_1 - x_2 \\
\text{subject to } &\quad x_1 + x_2 \geq 5 \\
&\quad 2x_1 - x_2 \leq 4 \\
&\quad x_1, x_2 \geq 0
\end{align*}
\]

Set up a Simplex tableau for the problem.

Solve and report (a) the optimal objective value, and (b) the optimal solution.
Part IV (30 points)

Find the intersection of the planes defined by

\[ 3x - 4y + 2z = 10 \]
\[ 2x - z = -2 \]
\[ -x - 4y + 4z = 14 \]

Give a geometric interpretation of the intersection of these planes.