BIOE 298, SECTIONS MFI & B

PRACTICE EXAM 2

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.
PART I (20 POINTS; 4 POINTS EACH)

(1) What is the distance between the origin and the hyperplane $3x_1 + 2x_2 - x_3 = 0$?

\[
\begin{align*}
\mathbf{a} \cdot \mathbf{x} = b & \implies \mathbf{a} \cdot \mathbf{x} = d \implies d = \frac{b}{\| \mathbf{a} \|} \\
\text{since } b = 0, d = \frac{0}{\| \mathbf{a} \|} = 0.
\end{align*}
\]

The hyperplane goes through the origin.

(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.)

The intersection is convex.

(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?

\[
\begin{align*}
\text{odds} &= \frac{P}{1-P} = \frac{1}{300} \implies 300P = 1-P \\
301P &= 1 \\
\implies P &= \frac{1}{301}
\end{align*}
\]

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

The matrix must have full row rank (which is equivalent to full rank for a square matrix).

(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.

Categorical outputs require logistic reg.

Categorical inputs do not

\(\implies\) example: the ECM HW problem.
You are given a set of inputs ($x$) and responses ($y$)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.1</td>
</tr>
<tr>
<td>1</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>5.8</td>
</tr>
<tr>
<td>3</td>
<td>11.3</td>
</tr>
</tbody>
</table>

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 $y = X\beta + \epsilon$.)

For $y = \beta_0 + \beta_1 x$

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

For $y = \beta_0 e^x$

$$X = \begin{bmatrix} e^0 \\ e^1 \\ e^2 \\ e^3 \end{bmatrix}$$
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>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>
</tr>
<tr>
<td>x</td>
<td>-0.345</td>
<td>0.385</td>
<td>-0.8961</td>
</tr>
<tr>
<td>x^2</td>
<td>1.125</td>
<td>0.12298</td>
<td>9.1476</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>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>
</tr>
<tr>
<td>expx</td>
<td>0.47699</td>
<td>0.023746</td>
<td>20.087</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?

*The first model has lower RMSE. It gives better predictions.*
Convert the following linear program to standard form:

$$\text{max } f = 2x_1 - x_2$$
subject to

$$x_1 + x_2 \geq 5$$
$$2x_1 - x_2 \leq 4$$
$$x_1, x_2 \geq 0$$

Set up a Simplex tableau for the problem:

$$\begin{bmatrix}
-1 & -1 & 0 & 0 & -5 \\
2 & -1 & 0 & 0 & 4 \\
-2 & 1 & 0 & 0 & 0
\end{bmatrix}$$

Solve and report (a) the optimal objective value, and (b) the optimal solution.

(a) Objective $f = 4$
(b) Solution

$x_1 = 2$
$x_2 = 0$
$s_1 = -3$
$s_2 = 0$
Find the intersection of the planes defined by
\[\begin{align*}
3x - 4y + 2z &= 10 \\
2x - z &= -2 \\
x - 4y + 4z &= 14
\end{align*}\]
Give a geometric interpretation of the intersection of these planes.

We don't find a specific solution using one \(\alpha\); the intersection is the set of all points at all \(\alpha\).

The third plane must contain the line formed by the intersection of the other two.