BIOE 210, SPRING 2019

EXAM 3

You have 80 minutes to complete this exam.
You may use notes or printouts from the course website,
but **no electronic resources**.
A standalone scientific or graphing calculator is allowed.

Circle your final answer for each question.

Name

Solutions

PART I (40 POINTS; 4 POINTS EACH)

(1) You fit a logistic regression model to predict if an individual will experience a heart attack in the next year given the number of times they exercise each week. The underlying linear model is

$$-3.99 - 0.105x$$

where x is the number of exercise sessions per week. What are the odds that someone who doesn't exercise at all will have a heart attack this year?

$$\text{odds}(x=0) = e^{-3.99} \approx 0.185$$

(2) How would the odds of a heart attack change if an individual who doesn't exercise started exercising once per week?

$$OR(\text{heart attack} \mid \text{exercise}+1) = e^{-0.105}$$

$$\approx 0.90$$

(3) If you fit a linear regression model with only an intercept term, the intercept will be

(a) The mean of all the input variables.
(b) The median of all the input variables.
(c) The mean of the response variable.
(d) The mean of a convex combination of all input variables.

(4) The matrix $\mathbf{A}$ has m rows and n columns with $m > n$. Which of the following is true?

(a) The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ has **more** nonzero eigenvalues than the matrix $\mathbf{A}\mathbf{A}^{\mathrm{T}}$.

(b) The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ has **the same number of** nonzero eigenvalues as the matrix $\mathbf{A}\mathbf{A}^{\mathrm{T}}$.

(c) The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ has **fewer** nonzero eigenvalues than the matrix $\mathbf{A}\mathbf{A}^{\mathrm{T}}$.

(d) There is not enough information to answer this problem.

The # nonzero eigenvalues for both is equal to the # of singular values of $\underline{X}$ $\Rightarrow$ they are the same.

(5) Write the inverse of the translation matrix $\begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{pmatrix}$.

$$\underline{T}(\Delta x, \Delta y)^{-1} = \underline{T}(-\Delta x, -\Delta y)$$

$$= \begin{pmatrix} 1 & 0 & -\Delta x \\ 0 & 1 & -\Delta y \\ 0 & 0 & 1 \end{pmatrix}$$

(6) Let the vector $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$ with $a > 0$ and $b > 0$. Compute the following:

(a) $\mathbf{R}(-180^{\circ})\mathbf{x}$

$$\underline{R}(-180^{\circ}) = \begin{pmatrix} \cos -\pi & -\sin -\pi \\ \sin -\pi & \cos -\pi \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\Rightarrow \underline{R}(-180^{\circ})\underline{x} = \boxed{\begin{pmatrix} -a \\ -b \end{pmatrix}}$$

(b) $\mathbf{R}(135^{\circ})^{-1}\mathbf{R}(270^{\circ})\mathbf{R}(45^{\circ})\mathbf{x}$

$$\underline{R}(135^{\circ})^{-1}\underline{R}(270^{\circ})R(45^{\circ})$$

$$= \underline{R}(-135^{\circ} + 270^{\circ} + 45^{\circ})$$

$$= \underline{R}(180^{\circ}) = \underline{R}(-180^{\circ})$$ [same as above]

$$\boxed{\begin{pmatrix} -a \\ -b \end{pmatrix}}$$

(7) Consider a standard die – a small cube with a number 1-6 written on each side. If you roll the die once.

(a) What are the odds you roll a 3?

$$P(3) = 1/6$$

$$\text{odds}(3) = \frac{P(3)}{1-P(3)} = \frac{1/6}{1-1/6} = 0.2 : 1$$

(b) What are the odds you roll an even number?

$$P(\text{even}) = 3/6$$

$$\text{odds}(\text{even}) = \frac{3/6}{1-3/6} = 1:1$$

(8) Which of the following linear models should include an intercept term? (Circle all that are correct.)

(a) A model that predicts a household's water usage based on the number of children in the house.

(b) A model that predicts insulin sensitivity based on serum concentration of an experimental drug.

(c) A model that predicts intelligence based on undergraduate GPA.

(d) A model that predicts intelligence based on the z-score of undergraduate GPA. (The z-score zero-centers a variable and normalizes it by the standard deviation.)

The intercept means:

(a) water usage in house with no children

(b) insulin sensitivity when no drug given

(c) Intelligence of student with 0.0 GPA

(d) Intelligence of student with average GPA.

(9) Circle TRUE or FALSE for the following questions.

(a) TRUE or FALSE. If an interaction term in a model is significant, the corresponding linear terms are also significant. See problem IV.

(b) TRUE or FALSE. Linear regression can be used to fit the model

$$y = \beta_1 \cos(x_1) + e^{\beta_2 x_2}$$ $\log y = \log \beta_1 + \log \cos x_1 + \beta_2 x_2$

(c) TRUE or FALSE. Unlike logistic regression, linear regression assumes the predictor variables are continuous. Predictors can be cont. or disc. for both.

(d) TRUE or FALSE. The preferred method for calculating the pseudoinverse of $\mathbf{X}$ is the formula $\mathbf{X}^+ = (\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}$.

Use SVD instead

(10) During Newton's method you arrive at a point $\mathbf{x}_i$.

(a) What should you do if $\mathbf{f}(\mathbf{x}_i) = \mathbf{0}$?

Stop; you're at a root.

(b) What should you do if $\mathbf{J}(\mathbf{x}_i) = \mathbf{0}$?

Pick a new initial guess.

Part II (15 points)

You measured the number of cells in a tissue culture flask each day for three days.

t [days]	$N(t)$ [cells]
0	12
1	29
2	102
3	336

You hypothesize that the cells are growing exponentially, i.e. $N(t) = N_0 e^{\mu t}$.

(a) Write a linear model you could fit to find values for N_0 and μ.

$$\log N = \underbrace{\log N_0}_{\beta_0} + \underbrace{\mu}_{\beta_1} t$$

(b) Set up a design matrix using your linear model and the data in the above table. You do not need to simplify your answers – expressions like e^{10}, $\log 4$, etc. are fine.

$$\underline{X} = (\underline{1} \;\; \underline{t}) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

(c) You fit your model and find that $\beta_0 = 2.39$ and $\beta_1 = 1.13$. What are the estimated values for N_0 and μ?

$$\log N_0 = \beta_0 \Rightarrow N_0 = e^{2.39}$$

$$\approx 10.9$$

$$\mu = \beta_1 = 1.13$$

PART III (15 POINTS)

Your goal is to find a point $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ such that $x_1 x_2 = 8$ and $-3x_1 + 8 = -x_2^2$.

(a) Rewrite this problem as a multivariable function $\mathbf{f}(\mathbf{x})$ that can be solved by root finding.

$$\underline{f}(\underline{x}) = \begin{pmatrix} x_1 x_2 - 8 \\ -3x_1 + 8 + x_2^2 \end{pmatrix} = \underline{0}$$

(b) Calculate Jacobian of $\mathbf{f}$ and evaluate it at $\mathbf{x} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$.

$$\underline{J}(\underline{x}) = \begin{pmatrix} x_2 & x_1 \\ -3 & 2x_2 \end{pmatrix}$$

$$\underline{J}\left(\begin{pmatrix} 4 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 2 & 4 \\ -3 & 4 \end{pmatrix}$$

(c) Is the point $\mathbf{x} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ a root of $\mathbf{f}$? Why or why not?

$$\underline{f}\left(\begin{pmatrix} 4 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 2 \cdot 4 - 8 \\ -3(4) + 8 + 2^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Since $\underline{f}(\underline{x}) = \underline{0}$, $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ is a root.

Part IV (10 points)

You collect a set of data describing the amount of **growth inhibition** of a tumor based on the dose of two drugs, A and B. You fit a linear model with an interaction term for both drugs, with the following results.

```
Linear regression model:
   inhibition ~ 1 + concA*concB

Estimated Coefficients:
                    Estimate       SE        tStat         pValue
                    --------    -------     ------     ----------

    (Intercept)     -0.84009    0.71872     -1.308        0.21755
    concA             2.9099     1.0626     3.4576      0.0001814
    concB               1.34    0.96668      1.235        0.35531
    concA:concB      -7.2842     1.4631     8.4544    4.5414e-04


Number of observations: 25, Error degrees of freedom: 21
Root Mean Squared Error: 0.6509
R-squared: 0.984, Adjusted R-Squared 0.982
F-statistic vs. constant model: 588, p-value = 1.23e-11
```

(a) Do both drugs A and B work alone to inhibit tumor growth? Use data from the modeling results to support your claim.

No - only A has both a positive coefficient and a significant p-value.

(b) Is there synergy between drugs A and B, i.e. is the combination of the two drugs better at inhibiting tumor growth? How do you know this from the linear modeling results?

No - the drugs show significant antagonism since the have a low p-value and a negative coefficient.

Part V (20 points)

The questions on this page refer to the MATLAB *output on the final page. Feel free to detach the final page when answering these questions.*

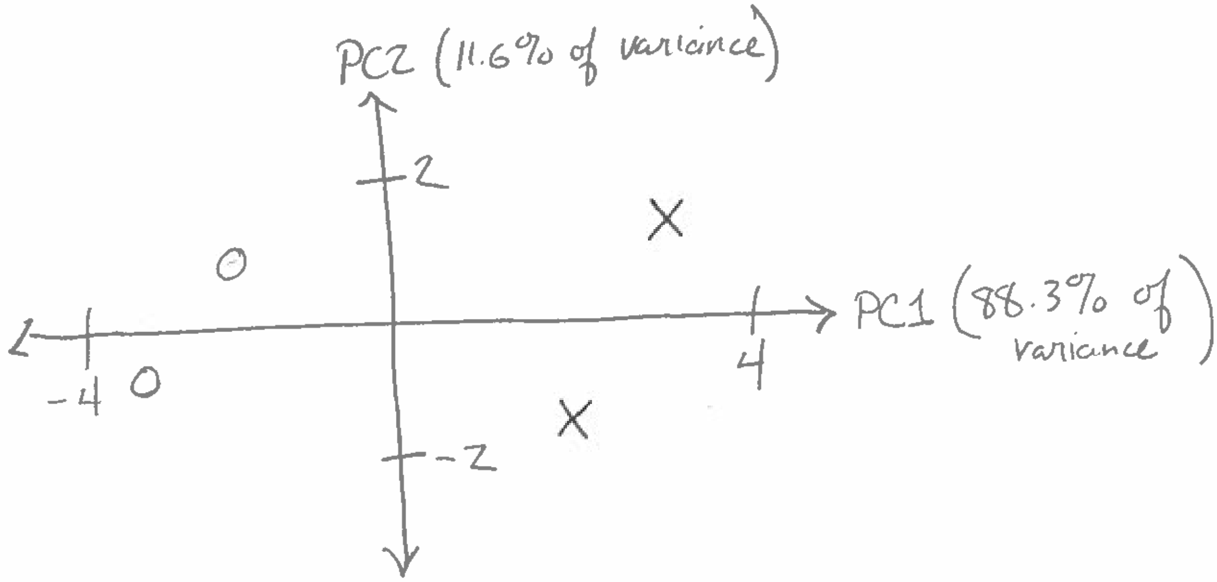
You measured the relative activity of four signaling proteins (ERK, MEK, Raf, and Akt) in a cancer cell line. The first two observations are taken from untreated cells. The second two observations are measured after incubation with a drug.

	ERK	MEK	Raf	Akt
Untreated 1	1.8147	0.0906	0.1270	-3.9134
Untreated 2	2.6324	0.0975	0.0278	-2.5469
Treated 1	-1.9575	0.0965	0.1576	0.9706
Treated 2	0.0957	0.4854	0.0800	3.1419

(a) Plot the four points using the first two principal components. Use circles for the untreated cells and X's for the treated cells. Label the percent of total variation explained by each principal component.

data are already in correct form (rows = samples, columns = variables), so use pca(data).

Plotting score(:,1) and score(:,2)



(b) Do either or both of these principal components separate the treated and untreated cells?

only PC1 separates treated from untreated.

(c) Which signaling proteins are **not** involved in the biological pathway targeted by the treatment?

only variables loaded onto PC1 are involved, so MEK + RAF are <u>not</u>.