TAKE-HOME CLASS QUIZ: DUE WEDNESDAY DECEMBER 4: ORDINARY LEAST SQUARES REGRESSION

MATH 196, SECTION 57 (VIPUL NAIK)

Your name (print clearly in capital letters): ______________________________________

PLEASE FEEL FREE TO DISCUSS *ALL* QUESTIONS.

(1) Assume no measurement error. Consider the situation where we have a function f of the form $f(x) = a_0 + a_1x$ with unknown values of the parameters a_0 and a_1. We collect n distinct input-output pairs, i.e., we collect n distonct inputs and compute the outputs for them. The coefficient matrix for the system is a $n \times 2$ matrix (the rows correspond to the input values, and the columns correspond to the unknown parameters). What is the rank of this matrix?
(A) It is always 2
(B) It is always n
(C) It is always $\min\{2, n\}$
(D) It is always $\max\{2, n\}$

Your answer: ____________________________

(2) Assume no measurement error. Consider the situation where we have a function f of the form $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$ with unknown values of the parameters a_0, a_1, $\ldots$, a_m. We collect n distinct input-output pairs, i.e., we collect n distinct inputs and compute the outputs for them. The coefficient matrix for the system is a $n \times (m+1)$ matrix (the rows correspond to the input values, and the columns correspond to the unknown parameters). What is the rank of this matrix?
(A) It is always $m + 1$
(B) It is always n
(C) It is always $\min\{m + 1, n\}$
(D) It is always $\max\{m + 1, n\}$

Your answer: ____________________________

(3) Assume no measurement error. Consider the situation where we have a function f of the form $f(x, y) = a_0 + a_1x + a_2y$ with unknown values of the parameters a_0, a_1, and a_2. We collect n distinct input-output pairs, i.e., we collect n distinct inputs (here an input specification involves specifying both the x-value and the y-value) and compute the outputs for them. The coefficient matrix for the system is a $n \times 3$ matrix (the rows correspond to the input values, and the columns correspond to the unknown parameters). What is the rank of this matrix?
(A) It is always $\min\{3, n\}$
(B) It is always $\max\{3, n\}$
(C) For $n = 1$, it is 1. For $n \geq 2$, it is 2 if the input points are all collinear in the xy-plane. Otherwise, it is 3.
(D) For $n = 1$, it is 1. For $n \geq 2$, it is 3 if the input points are all collinear in the xy-plane. Otherwise, it is 2.

Your answer: ____________________________

(4) Which of the following is closest to correct in the setting where we use a linear system to find the parameters using input-output pairs given a functional form that is linear in the parameters? Assume for simplicity that we are dealing with a functional form $y = f(x)$ with one input and one output, but possibly multiple parameters in the general description.

(A) The solutions to the linear system that we set up correspond to possibilities for the inputs to the function, and geometrically correspond to choices of points x for the graph $y = f(x)$.
(B) The solutions to the linear system that we set up correspond to possibilities for the inputs to the function, and geometrically correspond to different possible choices for the line or curve that is the graph $y = f(x)$.
(C) The solutions to the linear system that we set up correspond to possibilities for the parameters, and geometrically correspond to choices of points x for the graph $y = f(x)$.
(D) The solutions to the linear system that we set up correspond to possibilities for the parameters, and geometrically correspond to different possible choices for the line or curve that is the graph $y = f(x)$.

Your answer: ____________________________

(5) Continuing with the notation and setup of the preceding question, consider the coefficient matrix of the linear system. This matrix defines a linear transformation from the vector space of possible parameter values to the vector space of the outputs of the function. What is the image of this linear transformation?
(A) The image is the set of possible output values for which the linear system is consistent, i.e., we can find *at least one* function f of the required functional form that fits all the input-output pairs with *no measurement error*.
(B) The image is the set of possible output values for which the linear system has *at most one solution*, i.e., the set of output valus for which we can find *at most one* function f of the required functional form that fits all the input-output pairs with *no measurement error*.

Your answer: ____________________________

(6) Consider the case of polynomial regression for a polynomial function of one variable, allowing for measurement error. We believe that a function has the form of a polynomial. We can tentatively choose a degree m for the polynomial we are trying to fit, and a value n for the number of distinct inputs for which we compute the corresponding outputs to obtain input-output pairs (i.e., data points). We will get a $n \times (m + 1)$ coefficient matrix. Which of the following correctly describes what we should try for?
(A) We should choose n and $m + 1$ to be exactly equal, so that we get a unique polynomial.
(B) We should choose n to be greater than $m + 1$, so that the system is guaranteed to be consistent and we can find the polynomial.
(C) We should choose n to be less than $m + 1$, so that the system is guaranteed to be consistent and we can find the polynomial.
(D) We should choose n to be greater than $m + 1$, so that the system is *not* guaranteed to be consistent, but we do have a unique solution after we project the output vector to a vector for which the system is consistent.
(E) We should choose n to be less than $m + 1$, so that the system is *not* guaranteed to be consistent, but we do have a unique solution after we project the output vector to a vector for which the system is consistent.

Your answer: ____________________________

(7) Consider the general situation of linear regression. Denote by X the coefficient matrix for the linear system (also called the design matrix). Denote by $\vec{\beta}$ the parameter vector that we are trying to solve for. Denote by $\vec{y}$ an observed output vector. The idea in ordinary least squares regression is to choose a suitable vector $\vec{\varepsilon}$ such that the linear system $X\vec{\beta} = \vec{y} - \vec{\varepsilon}$ can be solved for $\vec{\beta}$. Among the many possibilities that we can choose for $\vec{\varepsilon}$, what criterion do we use to select the appropriate choice? Recall that the *length* of a vector is the square root of the sum of squares of its coordinates.
(A) We choose $\vec{\varepsilon}$ to have the minimum length possible subject to the constraint that $X\vec{\beta} = \vec{y} - \vec{\varepsilon}$ has a solution.

(B) We choose $\vec{\varepsilon}$ such that the system $X\vec{\beta} = \vec{y} - \vec{\varepsilon}$ can be solved and such that the solution vector $\vec{\beta}$ has the minimum possible length (among all such choices of $\vec{\varepsilon}$).

(C) We choose $\vec{\varepsilon}$ such that the system $X\vec{\beta} = \vec{y} - \vec{\varepsilon}$ can be solved and such that the difference vector $\vec{y} - \vec{\varepsilon}$ has the minimum possible length (among all such choices of $\vec{\varepsilon}$).

Your answer: ____________________________

(8) We have data for the logarithm of annual per capita GDP for a country for the last 100 years. We want to see if this fits a polynomial model. The idea is to try to first fit a polynomial of degree 0 (i.e., per capita GDP remains constant), then fit a polynomial of degree ≤ 1 (i.e., per capita GDP grows or decays exponentially), then fit a polynomial of degree ≤ 2 (i.e., per capita GDP grows or decays as the exponential of a quadratic function), and so on.

What happens to the length of the error vector $\vec{\varepsilon}$ as we increase the degree of the polynomial that we are trying to fit?

(A) The error vector $\vec{\varepsilon}$ keeps getting smaller and smaller in length, with a probability of almost 1 that it keeps *strictly* decreasing in length at each step, until the error vector becomes $\vec{0}$ (which we expect will happen when we get to the stage of trying to fit the function using a polynomial of degree 99).

(B) The error vector $\vec{\varepsilon}$ keeps getting larger and larger in length, with a probability of almost 1 that it keeps *strictly* increasing in length at each step, until the error vector becomes $\vec{y}$ (which we expect will happen when we get to the stage of trying to fit the function using a polynomial of degree 99).

Your answer: ____________________________