140.644: Statistical Machine Learning **Name:** _____

Instructor: Yiqun Chen TA: Junwei (Ivy) Sun

Quiz 3 — Practice — Classification: Model Flexibility and KNN

Not graded. Bring questions to class.

This is a shortened sample so you know what to expect. The real quiz has **4 multiple choice** and **2 questions in Part B**, and is worth 100 points; this one has 2 and 2, and the matching question uses three panels rather than four. The style and the difficulty are the same. No calculator is needed.

Part A: Multiple Choice [20 points]

Circle the single best answer. **(10 points)** each.

1. A KNN classifier is fitted with $K = 1$ and evaluated on a separate *test* set drawn from the same population. Its test error is:

   (A) always zero

   (B) generally greater than zero, because a test point is not its own nearest neighbor

   (C) always equal to the training error

   (D) undefined, because $K = 1$ cannot be applied to new observations

2. Consider these comparisons among logistic regression, LDA, and QDA. Which statements are correct?

   (I) QDA gives each class its own covariance matrix, so its decision boundary can be curved rather than a straight line.

   (II) LDA and logistic regression with linear terms both produce linear decision boundaries.

   (III) Logistic regression assumes the predictors are normally distributed within each class.

   (IV) QDA estimates fewer parameters than LDA, because it does not require a shared covariance matrix.

   (A) I only

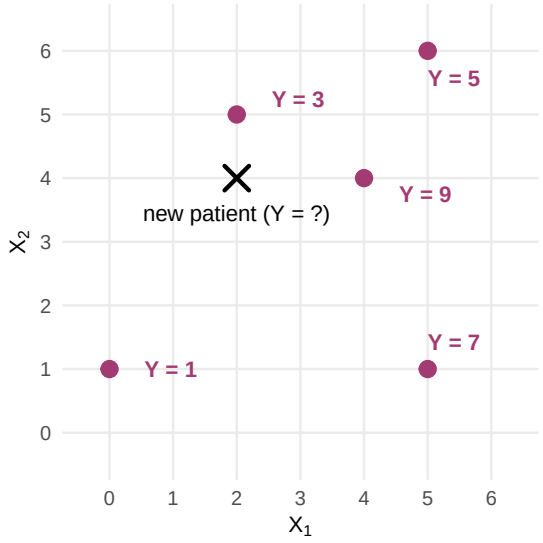
   (B) I and II only

   (C) II and III only

   (D) I, II, and IV only

Part B: K-Nearest Neighbors and Decision Boundaries [40 points]

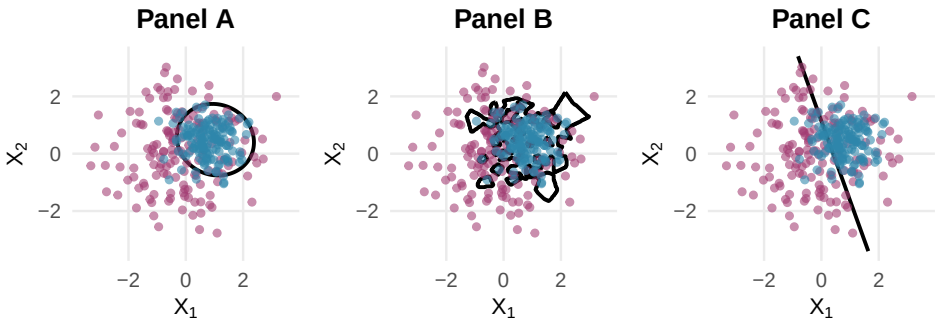
1. **(20 points)** The plot below shows five training observations, each drawn at its predictor values (X_1, X_2) and labeled with its observed response Y. The $\times$ marks a new patient. We predict the new patient's response with KNN regression, which averages Y over the K training observations closest to the new patient.



   (a) **(10 points)** With $K = 1$, the predicted response for the new patient is _____.

   (b) **(10 points)** With $K = 2$, the predicted response for the new patient is _____.

2. **(20 points)** The panels below show three classifiers fitted to the *same data*. The black curve is the estimated decision boundary. Each of the following models was used exactly once: LDA, QDA, and KNN with $K = 1$. Match each panel to its fitted model.



Panel A	Panel B	Panel C