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

Instructor: Yiqun Chen TA: Junwei (Ivy) Sun

Quiz 1 — Practice — Statistical Learning Foundations

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 **3 questions in Part B**, and is worth 100 points; this one has 2 and 2. The style, the difficulty, and the point values per question are the same. No calculator is needed. You may leave answers in terms of $\exp(\cdot)$ or $\log(\cdot)$.

Part A: Multiple Choice [20 points]

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

1. A model is fit on a training set and then evaluated on a separate test set drawn from the same population. In general we expect:
   (A) test error to be lower than training error, because the test set is smaller
   (B) training error to be lower than test error, because the model was fit to minimize it
   (C) the two to be equal, provided the split was random
   (D) training error to be lower only when the model is linear

2. A hospital wants to predict, for each admitted patient, the number of days they will stay. This is:
   (A) a classification problem, because patients are admitted or not
   (B) a regression problem, because the outcome is a number
   (C) an unsupervised problem, because no labels exist
   (D) none of the above

Part B: Reading an Analysis [40 points]

A clinic models patient waiting time in minutes against the number of staff on duty and the number of appointments booked that day. The fitted model is

$$\widehat{\text{wait}} \;=\; 40.0 \;-\; 5.0 \times (\text{staff on duty}) \;+\; 0.5 \times (\text{appointments}),$$

which is the same information as the software output below.

```
Call: lm(wait_min ~ n_staff + appointments)

               Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)     40.0000      4.0000     10.00     < 2e-16 ***
n_staff         -5.0000      1.0000     -5.00     6.5e-07 ***
appointments     0.5000      0.1000      5.00     5.7e-07 ***

Multiple R-squared:  0.61     (on the data used to fit)
```

1. **(20 points)** Using this model, compute the predicted waiting time on a day with 4 staff on duty and 60 appointments booked. Then state how the prediction changes if 2 more staff are added, with appointments unchanged.

2. **(20 points)** The clinic manager reads the negative coefficient on `n_staff` and concludes: "hiring two more staff will cut the average wait by 10 minutes." Do you agree? Give your reasoning.