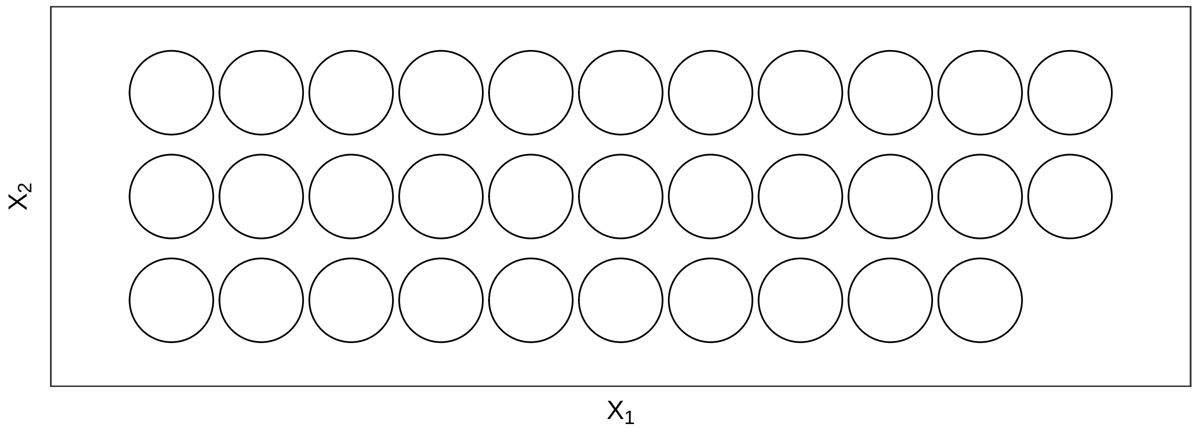
My arrival time: $y =$ ______ (relative to start of class, in decimal minutes rounded to nearest tenth).

The plot below represents the classroom as a predictor space on $X_1$ and $X_2$. Each circle represents the position of a potential observation. Record your arrival time in your circle as well as the arrival times of the 6 people nearest to you.




A $k$ nearest neighbor model predicts the response value at a given point $x$ as the average of the $y$ values of that point's $k$ nearest (in Euclidean distance) neighbors from the training data. That is:

$$\hat{f}(x) = \frac{1}{k} \sum_{x_i \in \mathcal{N}(x)} y_i$$

Focus on a point in the predictor space, $x$ (a two-element vector), that lies directly underneath your chair.

1. What is the $k = 1$ nearest neighbor prediction for this point?

2. What is the *squared training error*, $(y - \hat{f}(x))^2$, for this point? This is also knows as the *squared residual*.

3. Write these two values into the table on the following page, then repeat the process for $k \in \{3, 5, 7\}$. Once you have all of these values, add your squared residuals into the table on the board, then at the end, add the full training MSE to the table on the following page.

| | $k = 1$ | $k = 3$ | $k = 5$ | $k = 7$ |
|---|---|---|---|---|
| prediction | | | | |
| squared residual | | | | |
| full training MSE | | | | |

4. In the space below, draw a sketch for what the bias-variance tradeoff plot might look like for these models. Indicate each KNN model with a vertical line.

5. Select one of the empty seats in the room and label it $x_0$ on the previous page. Record the arrival times of its 6 nearest neighbors. Using what you suspect is the KNN model with the lowest *test MSE*, what is your prediction for that seat's arrival time?