1. Logistic regression assumes $p(y = 1|x) = \frac{1}{1 + \exp(-w^T x)}$. In class, we showed that if our prediction rule is “Predict class 1 if $p(y = 1|x) > 0.5$, otherwise predict 0”, the decision boundary of a logistic regression model with parameter $w$ is defined by $w^T x = 0$. Now if our prediction rule is “Predict class 1 if $p(y = 1|x) > 0.2$, otherwise predict 0” instead, what is the new decision boundary? In what situations would such a decision rule be practically useful?

   \[
P(y = 1|x) > 0.2 \Rightarrow \frac{1}{1 + \exp(-w^T x)} > 0.2 \Rightarrow 1 + \exp(-w^T x) < 5 \Rightarrow \exp(-w^T x) < 4 \Rightarrow -w^T x < \log_e 4 \Rightarrow w^T x > -\log_e 4
   \]

   This suggests that the new decision boundary is defined by $w^T x + \log_e 4 = 0$, which is more aggressive in predicting positive.

2. Define functional margin and geometric margin. Explain why functional margin is not a good objective to optimize in order to learn a maximum margin classifier.

   Solution:
   Functional margin of a point $X$ with respect to a linear decision boundary parameterized by $w$ and $b$ is defined as:
   \[y(w^T x + b)\]
   Geometric margin, on the other hand, is defined as the distance between $X$ and the decision boundary. Mathematically, it can be represented as:
   \[\frac{y(w^T x + b)}{||w||}\]

   The functional/geometric margin of a decision boundary is the smallest functional/geometric margin across all training examples. Functional margin is not a good objective to optimize because it can be scaled to be arbitrarily large without changing the decision boundary.
3. For the soft-margin SVM, parameter $c$ controls the trade-off between maximizing the margin and minimizing the slack variables (aka the ‘error’ of the fat decision boundary). Consider the following data set, what linear decision boundary will soft-margin SVM learn when $c=0.01$, and $c=100$? You don’t need to give specific equations, just mark out what the fat decision boundary look like in the figure.

Solution:
Consider the learning objective of SVM:

$$\text{argmin}_{\mathbf{w}, b} \frac{1}{2}||\mathbf{w}||^2 + c \sum_{i=1}^{n} \xi_i$$

subject to: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$

For $c=0.01$, $\xi_i$'s incur small penalties, thus it is ok ignoring some examples in order to achieve large margin. For $c=100$, on the other hand, $\xi_i$’s will cause large penalties. In this case, SVM will try to avoid to have any non-zero $\xi_i$'s.
4. Consider applying SVM to the following two class classification problem. Please
   a. Circle the support vectors (you should decide which points are support vectors by
      simply eyeballing)
   b. Mark out the fat decision boundary learned by SVM (this includes three lines
      \( w \cdot x + b = 0, \ w \cdot x + b = 1, \ w \cdot x + b = -1 \))
   c. What are the \( w \) and \( b \) parameter of this learned decision boundary (Hint: based
      on the support vectors, you should be able to plug their \( x \) values in the equations and
      solve for \( w \) and \( b \))

\[ \begin{align*}
3w_1 + 2w_2 + b &= 1 \quad \text{(1)} \\
5w_1 + 4w_2 + b &= -1 \quad \text{(2)} \\
6w_1 + 2w_2 + b &= -1 \quad \text{(3)}
\end{align*} \]

\( (3)-(1) = 3w_1 = -2 \Rightarrow w_1 = -2/3 \)

\( (2)-(3)*2 \Rightarrow -7w_1 - b = 1 \Rightarrow b = -7w_1 - 1 = 11/3 \)

\( \Rightarrow w_2 = -1/3 \)

\( \Rightarrow \text{decision boundary: } \frac{2}{3}x_1 - \frac{1}{3}x_2 + \frac{11}{3} = 0 \)