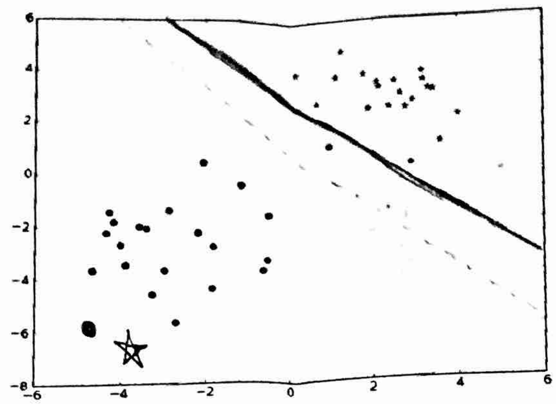
## 5 [16] Support Vector Machines

1. For this question, refer to the plot and dataset below, and recall that the soft-margin SVM optimization problem is equivalent to the following:

$$\min_{\mathbf{w},b} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^{n} \max \left[1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b), 0\right]$$



(a) (3) Suppose you train an SVM on the dataset above with a very *large* value for the regularization parameter $C$ (e.g., $C \to \infty$). Explain what would happen to the learned decision boundary as C increases, and draw as a bold solid line a decision boundary most likely to correspond to a very large value of $C$.

See solid line above.

(b) (3) Suppose that instead you were to use a *small* value of $C$. Explain what would happen to the learned decision boundary as $C$ *decreases*, and draw as a dashed line on the plot above a decision boundary most likely to correspond to a very small value of $C$.

See dashed line above.

(c) (1) On the plot, draw an additional *circle* data point that would not affect the decision boundary of a hard margin SVM trained on the dataset.

- Any point "left" of solid line and blue outlier.

(d) (1) On the plot, draw an additional *star* data point that would make the hard margin SVM optimization problem infeasible.

- Any point "left" of solid line and blue outlier.

2. Suppose you hand your dataset to the exciting new startup Bad Machine Learning Solutions, Inc., and they give you back a hard margin SVM classifier with parameters $\mathbf{w}$ and $b$. However, upon inspection, you discover that for every training data point $(\mathbf{x}_i, y_i)$, $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 2$.

(a) (1) Write down an equation for the separating hyperplane constraint used in the hard margin SVM problem.

$$\min_{w,b} w^T w \quad \text{s.t.} \quad y_i (w^T x_i + b) \geq 1$$

$|w^T x_i + b| \geq 1$ ok too

$\min\ y_i (w^T x_i + b) = 1$ ok too

$\min\ |w^T x_i + b| = 1$ ok too

(b) (2) Write down an equation for the margin of a linear classifier, $\gamma(\mathbf{w}, b)$.

$$\gamma(w, b) = \min_{x \in D} \frac{|w^T x + b|}{\|w\|_2^2}$$

-1 if missing numerator or denominator

(c) (5) Prove that the parameters this company gave you cannot possibly correspond to an optimal maximum margin classifier.

Solution 1:

$\hat{w} = w/2$ satisfies constraint

$$\gamma(\hat{w}, b) > \gamma(w, b)$$

$$\frac{|\hat{w}^T x_i + b|}{\|\hat{w}\|_2^2} > \frac{|w^T x_i + b|}{\|w\|_2^2}$$

$$\frac{|\frac{1}{2} w^T x_i + b|}{\|\frac{1}{2} w\|_2^2} > \frac{|w^T x_i + b|}{\|w\|_2^2}$$

$$\frac{|\frac{1}{2} w^T x_i + b|}{\frac{1}{4}\|w\|_2^2} > \frac{|w^T x_i + b|}{\|w\|_2^2}$$

$$\frac{4\,|\frac{1}{2} w^T x_i + b|}{\|w\|_2^2} > \frac{|w^T x_i + b|}{\|w\|_2^2} \quad ✓$$

$\hat{w}$ gives a larger margin, $w$ cannot possibly be the optimal parameter.

OR reasonable proof

-1 if small flaw or missing small step

-2 }
-3 } grader discretion based on correctness

-4 if effort...

-5 if empty