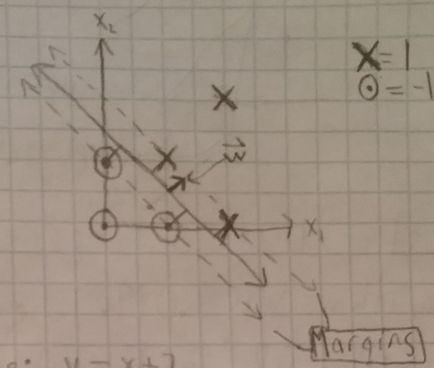
Meet Barot   Machine Learning Problem Set 2

1. a. Hyperplane:

$w_1 x_1 + w_2 x_2 + 3 = 0$

$\vec{w} = (-1,-1)$



Margin Calculations:

Margin from (0,2):

Intersecting line: $y = x+2$

Hyperplane: $y = 3-x$

Intersecting point: (0.5, 2.5)

Distance between (0.5, 2.5) and (0,2): $\sqrt{(2.5-2)^2+(.5)^2} = \frac{1}{\sqrt{2}}$

$2\gamma = \frac{2}{\sqrt{2}} = \sqrt{2}$

b. Take one point each from the two sets of points below, and you'll have a pair of support vectors.

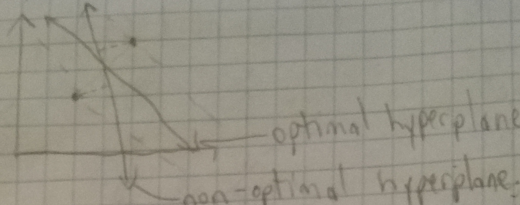
S1	S2
(4,0)	(2,0)
(2,2)	(0,2)

Ex: (4,0) and (2,0)

If you remove either set from the graph, the optimal hyperplane would change. This is why they are support vectors; the hyperplane can be constructed from them without knowing any other data points.

2. Given a data set of two n-dimensional points, the maximum margin hyperplane is one that is orthogo to the line that connects the two points. The hyperplane intersects the midpoint of this line, and so the margin γ is half the length of the line segment that connects the points. The hyperpl has to be orthogonal to this line, otherwise the distance of the poi

to the hyperplane would be less, and would not be the optimal.



3. The lowest test set errors for this ($\lambda = 2^{-9}$) SVM algorithm was .019. The lowest for the perceptron was .02, and for the averaged perceptron was .018. The support vectors are those data points such that $y_i(w_i \cdot x_i) \leq 1$. There are 180 for $\lambda = 2^{-8}$.

4. a. $\operatorname{sign}(\vec{w} \cdot \vec{x} + b) = \underset{k}{\operatorname{argmax}}(\vec{w}^{(1)} \cdot \vec{x} + b^{(1)}, \vec{w}^{(2)} \cdot \vec{x} + b^{(2)})$

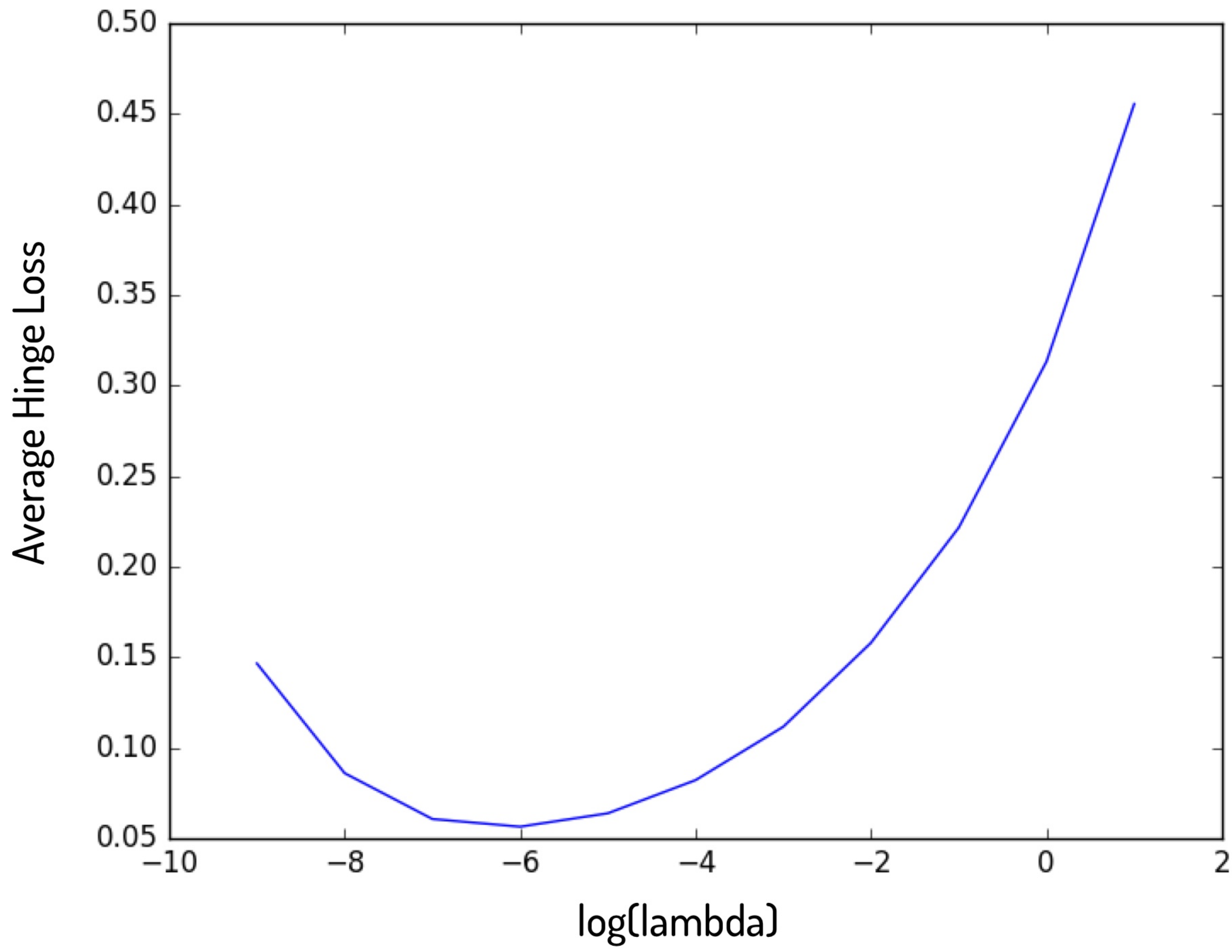
$\vec{w}^{(1)} x + b^{(1)} \geq \vec{w}^{(2)} x + b^{(2)}$ from the constraint.

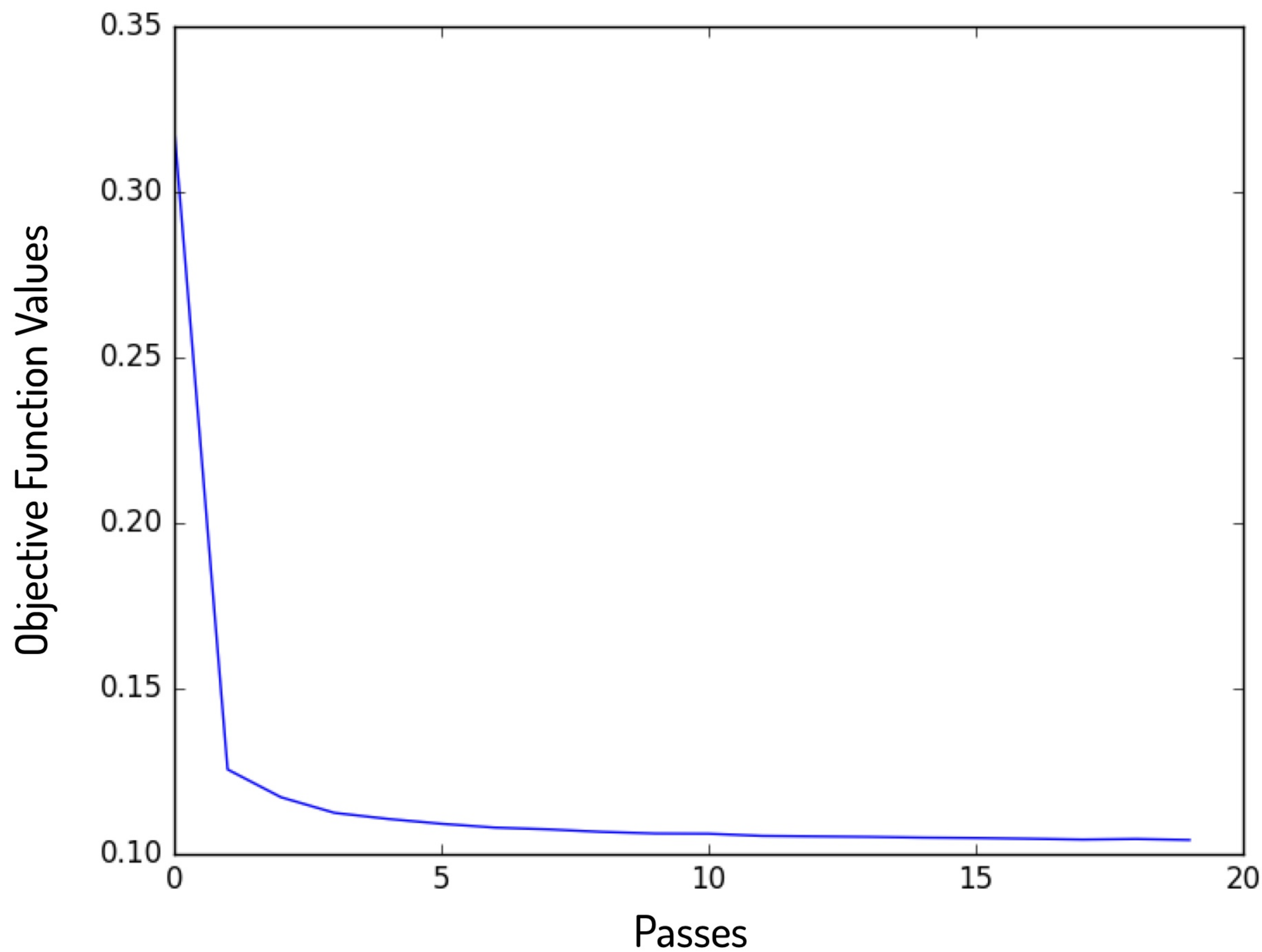
$\vec{w}^{(1)} x - \vec{w}^{(2)} x + b^{(1)} - b^{(2)} \geq 0$

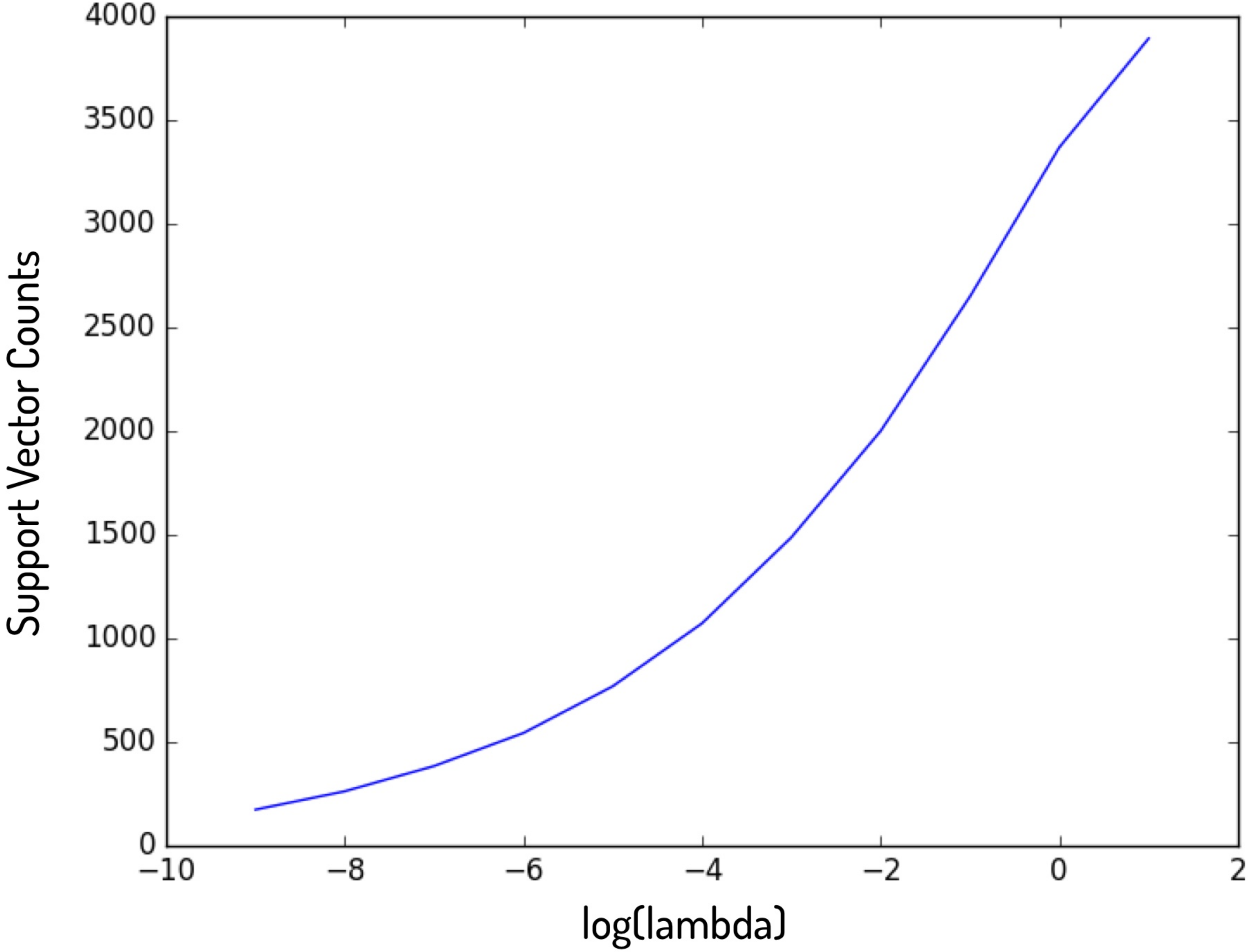
$(\vec{w}^{(1)} - \vec{w}^{(2)}) x + b^{(1)} - b^{(2)} \geq 0$ which is the same as $\vec{w} x + b \geq 0$. $\vec{w} = \vec{w}^{(1)} - \vec{w}^{(2)}$ and $b = b^{(1)} - b^{(2)}$.

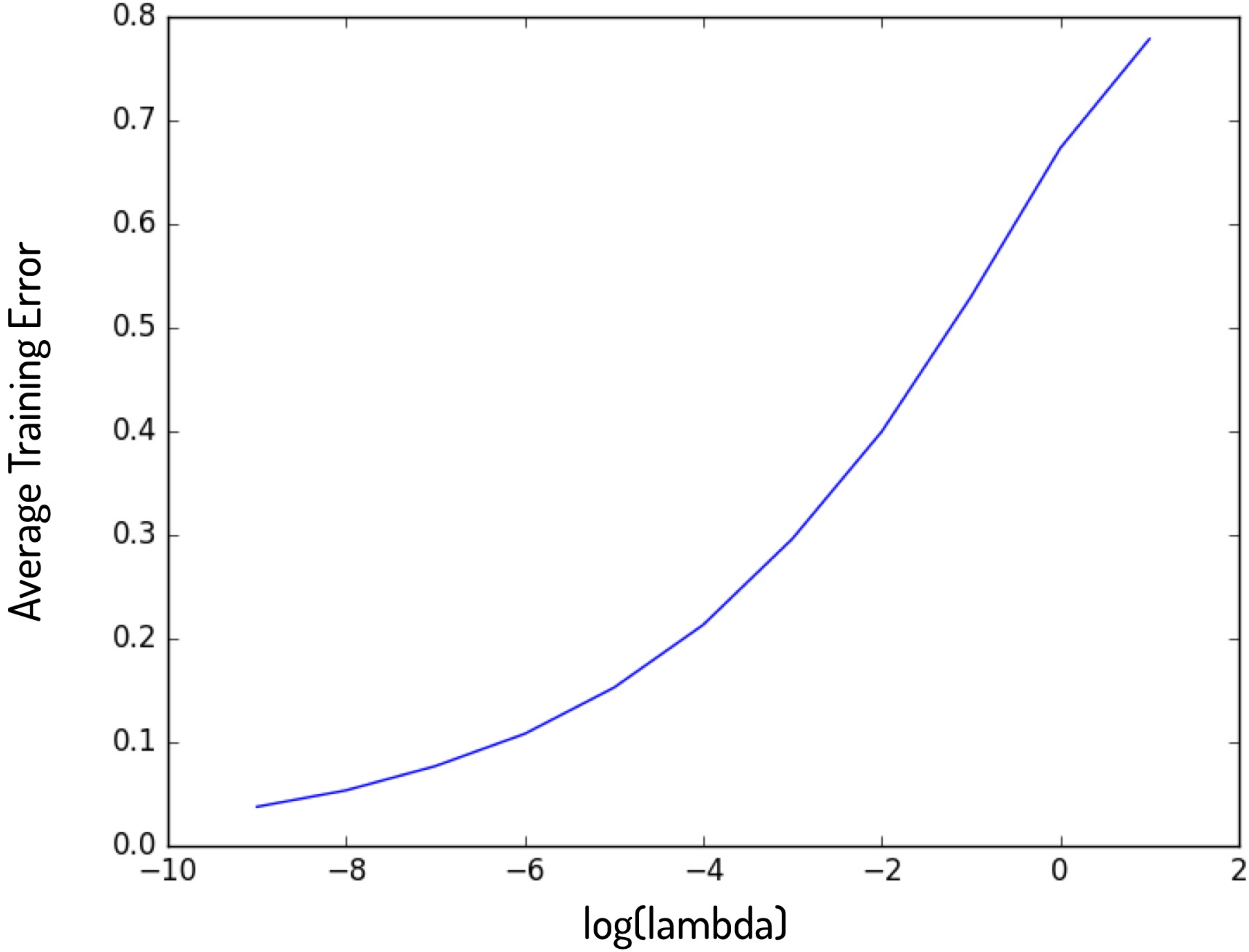
b. If we say that $\vec{w}^{(1)}$ and $\vec{w}^{(2)}$ are linearly independent, then $\vec{w} \cdot x + b \geq 0$ can be put into two inequalities: $\vec{w}^{(1)} \cdot x + b^{(1)} \geq 0$ and $\vec{w}^{(2)} \cdot x + b^{(2)} \geq 0$, which can be combined into $\vec{w}^{(1)} \cdot x - \vec{w}^{(2)} \cdot x + b^{(1)} - b^{(2)} \geq 0$. $(\vec{w}^{(1)} - \vec{w}^{(2)}) \cdot x + b^{(1)} - b^{(2)} \geq 0$.

Then, $\vec{w} = \vec{w}^{(1)} - \vec{w}^{(2)}$ and $b = b^{(1)} - b^{(2)}$.


0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
Average Hinge Loss
−10
−8
−6
−4
−2
0
2
log(lambda)


0.35
0.30
0.25
0.20
0.15
0.10
0
5
10
15
20
Objective Function Values
Passes


Support Vector Counts
4000
3500
3000
2500
2000
1500
1000
500
0
−10
−8
−6
−4
−2
0
2
log(lambda)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−10
−8
−6
−4
−2
0
2
Average Training Error
log(lambda)