assn4.1

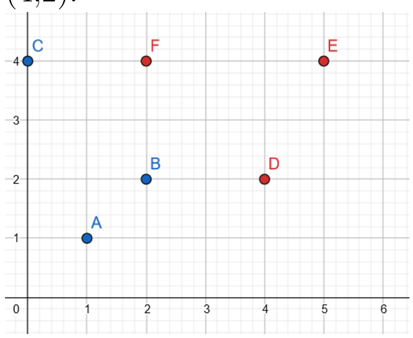
Daniel Loi
dtloi@ucsc.edu 1547401

Kevin Arellano
kcarella@ucsc.edu 1550271

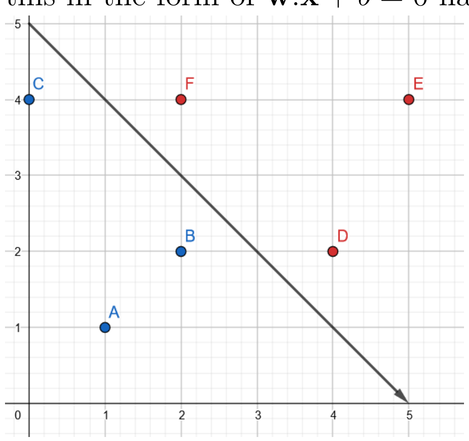
Claudio Sangeroki
csangero@ucsc.edu 1677403

October 2019

1. (a) Both classes are linearly separable. The support vectors are (0,4), (2,4), (2,2), (4,2).



(b) The max margin linear classifier is $y = -x + 5$ or $-x - y + 5 = 0$. We can write this in the form of $\mathbf{w}.\mathbf{x} + b = 0$ having $\mathbf{w} = (-1, -1)$, $\mathbf{x} = (x, y)$ and $b = 5$.



(c) The maximum margin is the distance from a hyperplane $\mathbf{w}.\mathbf{x} + b = 0$ to the closest point is given by $\frac{|\mathbf{w}.\mathbf{x}+b|}{||\mathbf{w}||}$. There are 4 equidistant points so the distance from the hyper plane to any of them is $\frac{1}{\sqrt{2}}$

(d) If we remove 1 support vector from the dataset, the maximum margin does not change. This is because there is another set of points that restrains the maximum margin.

(e) If a point that is not a support vector is removed from the dataset, then the decision boundary does not change. This is because the decision boundary is determined by support vectors.

2. (a) Since these points are correctly classified, then ξ_i will be 0.

   (b) Since these points do not lie on the decision boundary, then we will discuss points that are not far enough from the decision boundary and points that are mis-classified. If these points are not far enough from the decision boundary, then $0 < \xi_i < 1$. If these points are misclassified, then $\xi_i \geq 1$.

   (c) ——————————————————

   (d) A misclassified point has $\xi_i \geq 1$. To prove the equation $\#mistakes \leq \sum_i^n \xi_i$ we can look at two extreme scenarios. If all points n are correctly classified then $\#mistakes = 0$ and $\sum i^n \xi_i = 0$ which satisfies the equation $\#mistakes \leq \sum_i^n \xi_i$. If all points n are misclassified then $\#mistakes = n$ and ξ_i for every point is at least 1, meaning that $\sum_i^n \xi_i \geq n$ which also satisfies the equation $\#mistakes \leq \sum_i^n \xi_i$

   (e) We would expect a larger number of errors in C_1 because a lower C gives more importance to maximizing the margin and less importance on mistakes whereas a higher C will try harder avoid misclassification at the expense of making the margin smaller.

   (f) We can think of soft-margin SVM as hard-margin SVM for some value of C because soft-margin SVM is essentially hard-margin SVM if we had the ability to move mis-classified points to the correct side of the hyperplane. In order to account for the movement of the points, we add a Cost to each movement.

   (g) It is preferable to solve the dual optimization problem in this case.