Connor Malley

Support Vector Machines (SVMs)

Task 1:

Q1: What is the margin and support vectors?

Margin is the width a linear classifier could be increased before hitting any data points, and a support vector is a datapoint that the margin hits. (As defined in the powerpoints)

Q2: How does SVM deal with non-separable data?

An SVM deals with non-separable data by using a soft margin that allows for some misclassification error, or mapping the features to a higher dimensional space where they are separable by some hyperplane. You can also use a kernel to make the data linearly separable.

Q3: What is a kernel?

A kernel function calculates an inner product of the input features in order to make the data linearly separable.

Q4: How does a kernel relate to feature vectors?

The kernel applies an inner product to the feature vectors using some function such as a polynomial, gaussian, sigmoid, etc…

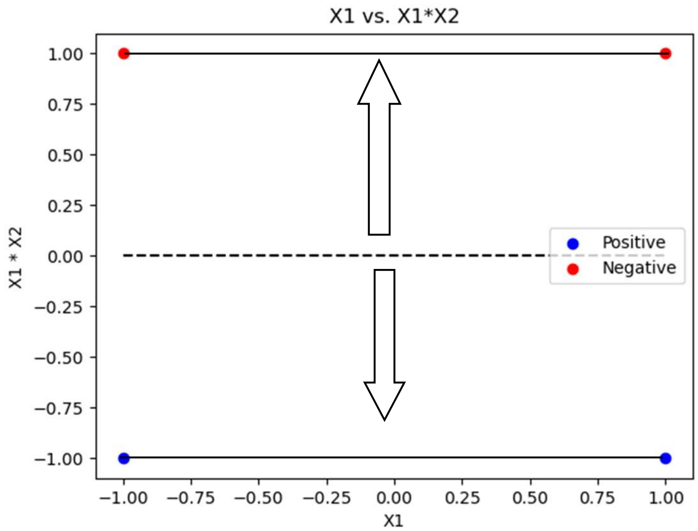
(I got these answers using the 5.3SVM2 powerpoint)

Task 2: Construct a support vector machine that computes the kernel function. Use four values of +1 and -1 for both inputs and outputs:

- **[−1, −1] (negative)**
- **[−1, +1] (positive)**
- **[+1, −1] (positive)**
- **[+1, +1] (negative).**

Map the input [x1, x2] into a space consisting of x1 and x1x2. Draw the four input points in this space, and the maximal margin separator. What is the margin?

➔ **1: X1 = -1, X2 = -1 -> X1 = -1, X1X2 = +1**
➔ **2: X1 = -1, X2 = +1 -> X1 = -1, X1X2 = -1**
➔ **3: X1 = +1, X2 = -1 -> X1 = +1, X1X2 = -1**
➔ **4: X1 = +1, X2 = +1 -> X1 = +1, X1X2 = +1**



→ **Margin = 1**

Task 3: Recall that the equation of the circle in the 2-dimensional plane is $(x_1 - a)^2 + (x_2 - b)^2 - r^2 = 0$. Please expand out the formula and show that every circular region is linearly separable from the rest of the plane in the feature space (x_1, x_2, x_1^2, x_2^2).

→ $(x1 - a)^2 + (x2 - b)^2 - r^2 = 0$
→ $X1^2 - 2a*x1 + a^2 + x2^2 - 2b*x2 + b^2 = 0$

→ $X1 = (X1^2 + a^2 + x2^2 - 2a*x2 + b^2) / 2a$
→ $X2 = (X1^2 + a^2 + x2^2 - 2a*x1 + b^2) / 2b$
→ $X1^2 = 2a*x1 - a^2 - x2^2 + 2a*x2 - b^2$
→ $X2^2 = 2a*x1 - a^2 - x1^2 + 2a*x2 - b^2$
→ **X1, X2, $X1^2$, and $X2^2$ are all linearly separable from the rest of the features**

Connor Malley

Task 4: Recall that the equation of an ellipse in the 2-dimensional plane is $c(x1 - a)^2 + d(x2 - b)^2 - 1 =$
Please show that an SVM using the polynomial kernel of degree 2, $K(u, v) = (1 + u \cdot v)^2$, is equivalent to a linear SVM in the feature space $(1, x1, x2, x1^2, x2^2, x1x2)$ and hence that SVMs with this kernel can separate any elliptic region from the rest of the plane.

Used powerpoints and "The Kernel Trick in Support Vector Classification" as a reference.

4)

$\bar{x}_i = [x_{i1} \quad x_{i2}]$ (original space)

Ellipse $\rightarrow$ $c(x_1-a)^2 + d(x_2-b)^2 - 1 = 0$

$$\Rightarrow \underset{\beta_3}{cx_1^2} - \underset{\beta_1}{2acx_1} + \underset{\beta_0}{ca^2} + \underset{\beta_4}{dx_2^2} - \underset{\beta_2}{2dbx_2} + \underset{\beta_0}{db^2} - \underset{\beta_0}{1} = 0$$

$$\Rightarrow (ca^2+db^2-1)1 - 2acx_1 - 2dbx_2 + cx_1^2 + dx_2^2 = 0$$

$$\Rightarrow \beta_0 \cdot 1 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1^2 + \beta_4 \cdot x_2^2 = 0$$

$$\Rightarrow \bar{x} = [1 \quad x_1 \quad x_2 \quad x_1^2 \quad x_2^2 \quad \overset{\beta = 0}{x_1 x_2}]$$ (feature space)

$$K(u,v) = (1+u \cdot v)^d = \left(1 + \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} [x_{j1} \quad x_{j2}]\right)^2$$

$$= (1 + (x_{i1}x_{j1}) + (x_{i2}x_{j2}))^2$$

$$= (1 + x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i1}x_{i2}x_{j1}x_{j2})$$

$$= [1 \quad \sqrt{2}x_{i1} \quad \sqrt{2}x_{i2} \quad \sqrt{2}x_{i1}x_{i2} \quad x_{i1}^2 \quad x_{i2}^2]^T \cdot [1 \quad \sqrt{2}x_{j1} \quad \sqrt{2}x_{j2} \quad \sqrt{2}x_{j1}x_{j2} \quad x_{j1}^2 \quad x_{j2}^2]$$

$$= \boxed{\varnothing(x_i)^T \varnothing(x_j)}$$ (in the feature space)

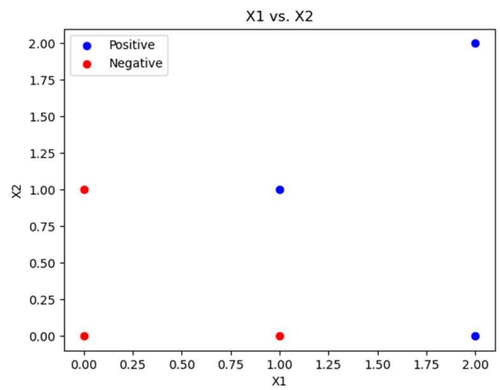
where $\varnothing(x) = [1 \quad \sqrt{2}x_1 \quad \sqrt{2}x_2 \quad x_1^2 \quad x_2^2 \quad \sqrt{2}x_1x_2]$

$\Rightarrow$ SVM using $K(u,v) = (1+u \cdot v)^2$ is equivalent to an SVM in the described feature space

Task 5: Consider the following training data

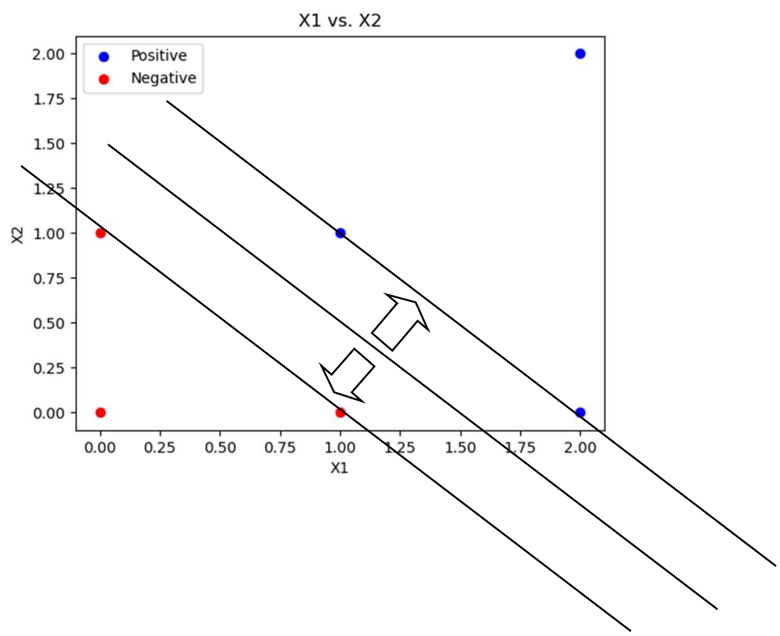
class	x_1	x_2
+	1	1
+	2	2
+	2	0
−	0	0
−	1	0
−	0	1

(a) Plot these six training points. Are the classes {+, −} linearly separable?



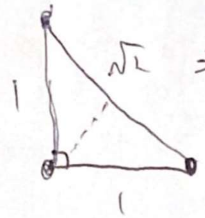
From the above plot , it is clear that the classes are linearly separable, (by a line with negative slope and y-intercept ~= 1.5)

b) Construct the weight vector of the maximum margin hyperplane by inspection and identify the support vectors.



5

class	x_1	x_2
+	1	1
+	2	2
+	2	0
−	0	0
−	1	0
−	0	1



$$\Rightarrow \left(\frac{M}{2}\right)^2 = 1^2 - \left(\frac{\sqrt{2}}{2}\right)^2$$

$$\Rightarrow \frac{M}{2} = \sqrt{1 - \frac{1}{2}}$$

$$\Rightarrow M = 1/\sqrt{2}$$

$\Rightarrow$ want to minimize $\Phi(w) = \frac{1}{2} w^t w$ w/ $y_i(w x_i + b) \geq 1$

where $\bar{w} = [\beta_1 \ \beta_2]$

support vectors are
$$\boxed{\begin{array}{lll} S_1 = [0 \ 1], & S_2 = [1 \ 0] & (-1) \\ S_3 = [1 \ 1], & S_4 = [2 \ 0] & (+1) \end{array}}$$

$$\Rightarrow M = \frac{2}{|w|} \Rightarrow M^2 = \frac{4}{w^T w} \Rightarrow w^T w = \frac{4}{M^2}$$

$$M = \frac{1}{\sqrt{2}} \Rightarrow w^T w = \frac{4}{\left(\frac{1}{2}\right)} = 8 = [\beta_1 \ \beta_2]\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix}$$

$$\Rightarrow \beta_1^2 + \beta_2^2 = 8 \Rightarrow \boxed{\beta_1 = \pm\sqrt{8 - \beta_2^2}}$$

Support vectors are those data points that define the hard margin. In this case there are 4: S_1 = {0, 1} (-), S_2 = {1, 0} (-), S_3 = {1, 1} (+), S_4 = {2, 0}.
There is also a margin of 1/(sqrt(2)) in this case.

Task 6: Consider a dataset with 3 points in 1-D:

(class)	x
$+$	0
$-$	-1
$-$	$+1$

(a) Are the classes {+, −} linearly separable?

The classes are not linearly separable, as the + class is between two points in the – class on the x-axis

(b) Consider mapping each point to 3-D using new feature vectors φ(x) = [1, sqrt(2)x, x^2] . Are the classes now linearly separable? If so, find a separating hyperplane.

See next page for work…

Connor Malley

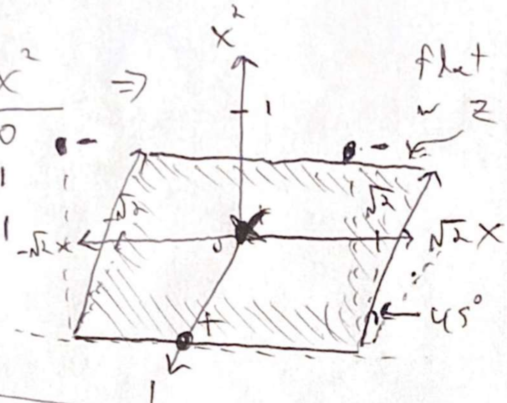
6b

let $\phi(x) = [1 \quad \sqrt{2}x \quad x^2]$

$\Rightarrow$

class	1	$\sqrt{2}x$	x^2
+	1	0	0
−	1	$-\sqrt{2}$	1
−	1	$+\sqrt{2}$	1

$\Rightarrow$



flat hyperplane w $z = 0.5$ gives max margin

$\Rightarrow$ By inspection, classes (+ and −) are linearly separable using $\phi(x) = [1 \quad \sqrt{2}x \quad x^2]$

$\Rightarrow$ They are all support vectors,

$\Rightarrow$ let $\beta_0 + \beta_1 \cdot (\sqrt{2}x) + \beta_2 x^2 = 0$ be the separating hyperplane

$\Rightarrow$ $\vec{w} = [\beta_0 \quad \beta_1 \quad \beta_2]$ $\Rightarrow$ mimize $\phi(w) = \dfrac{\beta_0^2 + \beta_1^2 + \beta_2^2}{2}$

$\Rightarrow$ separating hyperplane (w/ max margin) by inspection is:

(in feature space) $\Rightarrow$ $\boxed{(\underset{z}{x^2}) - 0.5 = 0}$

with margin $\boxed{M = 1}$

Task 7: Learning SVMs on the Titanic dataset. Please report your five-fold cross validation classification accuracies on Titanic training set, with respect to the linear, quadratic, and RBF kernels. Which kernel is the best in your cases?

Linear Kernel Accuracy (5-fold): 0.671156
Quadratic Kernel Accuracy (5-fold): 0.672278
RBF Kernel Accuracy (5-fold): 0.705948

In this case, the RBF kernel performed the best (~70% accuracy)
See code for reference.

GITHUB LINK:

https://github.com/malleyconnor/cap5610/tree/master/hw4