Homework 5 Arthur Bražinskas
11138904

N1

$$\mathcal{J}(a, R, \xi_i, \alpha_i, \mu_i) = R^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \underbrace{\left(-\|x_i - a\|^2 + R^2 + \xi_i\right)}_{\substack{g_i \\ \text{(constraint)}}} - \sum_{i=1}^{N} \mu_i \xi_i$$

N2

1) $\alpha_i \geq 0$

2) $g_i \geq 0 \Rightarrow -\|x_i - a\|^2 - R^2 + \xi_i \geq 0$

3) $\alpha_i g_i = 0$

4) $\mu_i \geq 0$

5) $\xi_i \geq 0$ For $i = 1 \ldots N$

6) $\mu_i \xi_i = 0$

In addition we seach for $\nabla f = \lambda \nabla g$, so

7) $\nabla_a \mathcal{J} = 0 \Rightarrow -\sum_{i=1}^{N} -2 \cdot \alpha_i (x_i - a)(-1) = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i x_i = \sum_{i=1}^{N} \alpha_i a \Rightarrow$

$$\Rightarrow a = \frac{\sum_{i=1}^{N} \alpha_i x_i}{\sum_{i=1}^{N} \alpha_i} \Rightarrow a = \sum_{i=1}^{N} \alpha_i x_i \quad \left(\text{as } \sum_{i=1}^{N} \alpha_i = 1\right)$$

8) $\nabla_{R^2} \mathcal{J} = 0 \Rightarrow 1 - \sum_{i=1}^{N} \alpha_i = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i = 1$

9) $\nabla_{\xi_i} \mathcal{J} = 0 \Rightarrow C - \alpha_i - \mu_i \Rightarrow \alpha_i = C - \mu_i$

N3

Complete slackness conditions are: 3 and 6 (in N2): $\alpha_i g_i = 0$ and $\mu_i \xi_i = 0$ for $i: 1 \ldots N$

1) if $\alpha_i > 0$ then g_i must be 0, so: $-\|x_i - a\|^2 + R^2 + \xi_i = 0 \Rightarrow$

$\Rightarrow \|x_i - a\|^2 = R^2 + \xi_i$, so the point x_i is either outside of the circle or on the border (SV), depending on ξ_i.

2) if $\mu_i > 0$ then $\xi_i = 0$, so x_i is inside of the circle or on the border (SV)

N4

$$\tilde{\mathcal{J}} = R^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i \left(-\Big\|x_i - \sum_{j=1}^{N} \alpha_j x_j\Big\|^2 + R^2 + \xi_i\right) - \sum_{i=1}^{N} \mu_i \xi_i =$$

$$= \cancel{R^2} + C\sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \alpha_i \Big\|x_i - \sum_{j=1}^{N} \alpha_j x_j\Big\|^2 - \cancel{\underbrace{\sum_{i=1}^{N} \alpha_i}_{=1} R^2} - \sum_{i=1}^{N} \alpha_i \xi_i - \sum_{i=1}^{N} \mu_i \xi_i =$$

$$= C \cancel{\sum_{i=1}^{N} \xi_i} + \sum_{i=1}^{N} \alpha_i \Big\|x_i - \sum_{j=1}^{N} \alpha_j x_j\Big\|^2 - \cancel{\sum_{i=1}^{N} \xi_i C} + \cancel{\sum_{i=1}^{N} \mu_i \xi_i} - \cancel{\sum_{i=1}^{N} \mu_i \xi_i} =$$

N4 (continued)

$$= \sum$$

$$\sum_{i=1}^{N} \|\theta_i\|^2 = \sum_{i=1}^{N} \theta_i^T \theta_i$$

$$= \sum_{i=1}^{N} \alpha_i \|x_i - \sum_{j=1}^{N} \alpha_j x_j\|^2 = \sum_{i=1}^{N} \alpha_i (x_i - \sum_{j=1}^{N} \alpha_j x_j)^T (x_i - \sum_{j=1}^{N} \alpha_j x_j) =$$

$$= \sum_{i=1}^{N} \alpha_i (x_i^T x_i - 2 x_i^T \sum_{j=1}^{N} \alpha_j x_j + \sum_{j=1}^{N} \alpha_j x_j^T \sum_{m=1}^{N} \alpha_m x_m) =$$

those become identical

$$= \sum_{i=1}^{N} \alpha_i \underbrace{x_i^T x_i}_{k(x_i,x_i)} - 2 \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \underbrace{x_i^T x_j}_{k(x_i,x_j)} + \underbrace{\sum_{i=1}^{N} \alpha_i}_{=1} \sum_{j=1}^{N}\sum_{m=1}^{N} \alpha_j \alpha_m \underbrace{x_j^T x_m}_{k(x_j,x_m)} =$$

kernals

$$= \sum_{i=1}^{N} \alpha_i k(x_i,x_i) - \sum_{j=1}^{N}\sum_{m=1}^{N} \alpha_j \alpha_m k(x_j,x_m) \Leftarrow$$ dual kernalized problem (objective func.)

N5

1) $\alpha_i = C - \mu_i \Rightarrow \alpha_i - C = -\mu_i \Rightarrow \mu_i = C - \alpha_i$

2) $a = \sum_{i=1}^{N} \alpha_i x_i$

3) if $\xi_i \neq 0$: $\varepsilon_i = \|x_i - a\|^2 - R^2$, and $\varepsilon_i > 0$ for outliers. (we shall replace a and R^2 in (5))

4) Our Radius depends on the support vectors, so we can continue the way of the eq. in (3) and get: $R^2 = \frac{1}{M} \sum_{i=1}^{M} \|x_i - \sum_{j=1}^{N} \alpha_j x_j\|^2$

Notice that our R is computed by taking an averaged sum over M support vectors, and we know that for S.V. $\varepsilon_i = 0$ for $i = 1 \dots M$

5) Now we rewrite (3) and say that $\varepsilon_i = 0$ if x_i is the S.V. or inside of the circle

or $\xi_i = \|x_i - \sum_{i=1}^{N} \alpha_i x_i\|^2 - \frac{1}{M} \sum_{j=1}^{M} \|x_j - \sum_{m=1}^{N} \alpha_m x_m\|^2 > 0$

The latter case of ξ_i is for x_i that are outliers.

N6

$y(x_i) = \|x_i - a\|^2 - R^2 > 0$ for outliers, now we dualize and kernalize

$$\|x_i - \sum_{j=1}^{N} \alpha_j x_j\|^2 - \frac{1}{M}\sum_{j=1}^{M} \|x_j - \sum_{m=1}^{N} \alpha_m x_m\|^2 > 0 \Rightarrow$$ (we apply the same trick $\sum_i \|\theta_i\|^2 = \sum_i \theta_i^T \theta_i$)

$$\Rightarrow \left(x_i - \sum_{j=1}^{N} \alpha_j x_j\right)^T \left(x_i - \sum_{j=1}^{N} \alpha_j x_j\right) > \frac{1}{M}\sum_{j=1}^{M}\left(x_j - \sum_{m=1}^{N} \alpha_m x_m\right)^T \left(x_j - \sum_{m=1}^{N} \alpha_m x_m\right) \Rightarrow$$

$$\Rightarrow \underbrace{x_i^T x_i}_{k(x_i,x_i)} - 2\sum_{j=1}^{N} \alpha_j \underbrace{x_i^T x_j}_{k(x_i,x_j)} + \sum_{j=1}^{N}\sum_{m=1}^{N} \alpha_j \alpha_m \underbrace{x_j^T x_m}_{k(x_j,x_m)} > \frac{1}{M}\sum_{j=1}^{M} \underbrace{x_j^T x_j}_{k_j} - \frac{2}{M}\sum_{j=1}^{M}\sum_{m=1}^{N} \alpha_m \underbrace{x_j^T x_m}_{k(x_j,x_m)} + \frac{M}{M}\sum_{j=1}^{N}\sum_{m=1}^{N} \alpha_j \alpha_m x_j x_m$$

$$\Rightarrow k(x_i,x_i) - 2\sum_{j=1}^{N} \alpha_j k(x_i,x_j) > \frac{1}{M}\left(\sum_{j=1}^{M}\left(k(x_j,x_j) - 2\sum_{m=1}^{N} \alpha_m k(x_j,x_m)\right)\right)$$

If it's not obvious, then we could think that if the equation is satisfied, then x_i is an outlier.

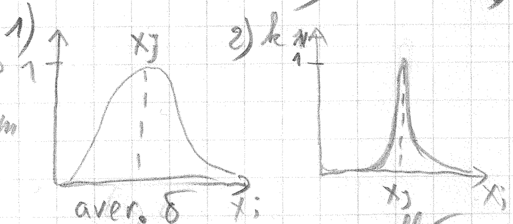
N7

1) if $C=0$, the cost for outliers is zero, so we will end up having small $R \to 0$ and a lot of outliers.

2) if $C=\infty$, the radius will cover merely every point, so that we will not have any outliers.

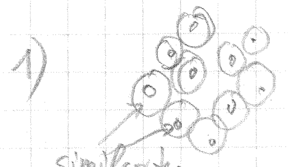
N8

$k(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\delta^2}\right)$, so this kernel measuring the similarity of two vectors will turn into a binary identifying ~~function~~ function if $\delta \to 0$.



aver. δ (cartoon sketches) small δ

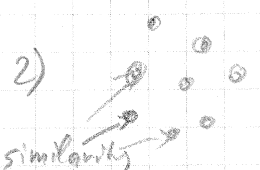
Geometrically in the context of the problem it will mean the following



similarity regions

of x_i $i=1..N$ (train points)

aver. δ



similarity regions

small δ

Thus if δ is too small it will lead to a lot of outliers during the test, unless test points are identical to the training data points.

N8 (continued)

The difference between linear kernal and gaussian kernal based solutions is that for x and x' pair with increase of x' the liner kernal outputs will increase linearly while the gaussian kernals with very small δ will return values close to 0 as x and x' start to differ.

N 9

We rewrite our objective function sightly and introduce a hinge function.

$$\min_{a,R} f(a,R) = R^2 + C\sum_{i=1}^{N} \max(0, -t_i y(x_i))$$

$$y(x_i) = \|x_i - a\|^2 - R^2$$

Intuition is the following:

1) if $y_i = y(x_i) = 0$, we have a support vector x_i and the cost is 0
2) if $y_i < 0$ - x_i is inlier.
3) if $y_i > 0$ - x_i is outlier.

E.g. $t_i = -1$ and $y_i = 1000$, then the cost is $C \cdot 1000$ as x_i should have been an inlier, but was classified as an outlier.