Introduction to Machine Learning - hw 3 - short

Omer Simhi, 316572593

May 15, 2021

1. For the (homogeneous) linearly separable case:

   (a) When $\lambda \to \infty$ to which solution will the soft SVM converge?
   Solution - Notice that:

   $$\lim_{\lambda\to\infty}\left(argmin_{w\in\mathbb{R}^d}\left(\frac{1}{m}\cdot\sum_{i=1}^{m}\max\left\{0,1-y_i w^T x_i\right\}+\lambda\left\|w\right\|_2^2\right)\right)\approx$$
   $$\approx\lim_{\lambda\to\infty}\left(argmin_{w\in\mathbb{R}^d}\left(\lambda\left\|w\right\|_2^2\right)\right)$$

   since the effect of $\frac{1}{m}\cdot\sum_{i=1}^{m}\max\left\{0,1-y_i w^T x_i\right\}$ when $\lambda\to\infty$ is negligible compare to $\lambda\left\|w\right\|_2^2$. Notice that $\lim_{\lambda\to\infty}\left(argmin_{w\in\mathbb{R}^d}\left(\lambda\left\|w\right\|_2^2\right)\right)<\infty$ iff $\left\|w\right\|_2^2=0$ otherwise $\lim_{\lambda\to\infty}\left(argmin_{w\in\mathbb{R}^d}\left(\lambda\left\|w\right\|_2^2\right)\right)=\infty$. So we get $\left\|w\right\|_2=0$ and so $w=0$ is the only way to minimize the expression.

   (b) When $\lambda\to 0$, the soft SVM converges to the hard SVM's solution. Explain briefly and intuitively how it can be seen from the formulations above.
   Solution - Recall that λ is the "tradoff factor" between increasing the margin size and ensuring that each of x_i are in correct size of the margin. Now, as $\lambda\to 0$ we get that the expression $\lambda\left\|w\right\|_2^2$ is negligible and thus the problem became : $argmin_{w\in\mathbb{R}^d}\left(\frac{1}{m}\cdot\sum_{i=1}^{m}\max\left\{0,1-y_i w^T x_i\right\}\right)$. Notice that this means that we can't have violations of margin constraints, and so the second term, $1-y_i w^T x_i$ is negligible so we returns to the hard SVM solution.

2. Let $K_1(u,v) = \langle\phi_1(u), \phi_1(v)\rangle$, $K_2(u,v) = \langle\phi_2(u), \phi_2(v)\rangle$ where $\phi_1 : \mathcal{X} \to \mathbb{R}^{n_1}, \phi_2 : \mathcal{X} \to \mathbb{R}^{n_2}$ s.t $n_1, n_2 \in \mathbb{N}$. Let's denote:

$$\phi_1(u) = \begin{pmatrix} u_1 & . & . & . & u_{n_1} \end{pmatrix}^T, \phi_1(v) = \begin{pmatrix} v_1 & . & . & . & v_{n_1} \end{pmatrix}^T$$
$$\phi_2(u) = \begin{pmatrix} u'_1 & . & . & . & u'_{n_2} \end{pmatrix}^T, \phi_1(v) = \begin{pmatrix} v'_1 & . & . & . & v'_{n_2} \end{pmatrix}^T$$

We get:

$$K_3(u,v) := K_1(u,v) + K_2(u,v) = \langle\phi_1(u), \phi_1(v)\rangle + \langle\phi_2(u), \phi_2(v)\rangle$$

$$= \sum_{i=1}^{n_1} u_i v_i + \sum_{j=1}^{n_2} u'_j v'_j = \begin{pmatrix} u_1 & . & . & . & u_{n_1} u'_1 ... u'_{n_2} \end{pmatrix} \begin{pmatrix} v_1 \\ . \\ . \\ v_{n_1} \\ v'_1 \\ . \\ . \\ . \\ . \\ v'_{n_2} \end{pmatrix} =$$

$$= \left\langle \begin{pmatrix} u_1 & . & . & . & u_{n_1} u'_1 ... u'_{n_2} \end{pmatrix}, \begin{pmatrix} v_1 & . & . & . & v_{n_1} v'_1 ... v'_{n_2} \end{pmatrix} \right\rangle = \langle\phi_1(u)\phi_2(u), \phi_1(v)\phi_2(v)\rangle$$

where $\phi_1(u)\phi_2(u)$ is the concatenation of $\phi_1(u)$ to $\phi_2(u)$. So, if we define $\phi_2 : \mathcal{X} \to \mathbb{R}^{n_3}$ with $n_3 := n_1 + n_2$ and:

$$\phi_3(x) = (\phi_1(x)\phi_2(x))^T$$

we get finally $K_3(u,v) = \langle\phi_3(u), \phi_3(v)\rangle$ as required.

3. Define the hypothesis class of axis aligned rectangles (or cuboids) in $\mathbb{R}^d$ -

$$\mathcal{X} = \mathbb{R}^d, \mathcal{H}^d_{rect} = \left\{ h_\theta \mid \forall i \in [d] : \theta^{(1)}_i < \theta^{(2)}_i \right\}$$

where $\theta^{(1)}, \theta^{(2)} \in \mathbb{R}^d$, $\theta = \left(\theta^{(1)}, \theta^{(2)}\right)$ and $h_\theta(x) = \begin{cases} 1 & \wedge_{i\in[d]} \left(\theta^{(1)}_1 \leq x_i \leq \theta^{(2)}_i\right) \\ -1 & otherwise \end{cases}$

(a) Explain in your own simple words (1-3 sentences), what do we need to show in order to prove that $VCdim\left(\mathcal{H}^d_{rect}\right) = k$ for some $k \in \mathbb{N}$.

Solution - We will show two things by defenition:

- There exist a sample set S with size k s.t $\exists h_\theta \in \mathcal{H}^d_{rect}$ that shatter S for any given labels set for the set S.
- For any sample set S with size $k+1$ there is no $h_\theta \in \mathcal{H}^d_{rect}$ that shatter S.

(b) Prove that $VCdim\left(\mathcal{H}^d_{rect}\right) \geq 2d$.

Solution - Let us consider the following set S:

$$S = \bigcup_{i=1}^{d} \left\{e^+_i, e^-_i\right\}$$

$$e^+_i := \left(0, 0, ..., \underbrace{1}_{i}, 0, ..., 0\right), e^-_i = \left(0, 0, ..., \underbrace{-1}_{i}, 0, ..., 0\right)$$

clearly $|S| = 2d$. Denote an arbitrary set of labels $Y := \{y_1, y'_1, ..., y_d, y'_d\}$ for the $2d$ points of S where y_i is the label of e^+_i and y'_i is the label of e^-_i. Now, for $i = 1, 2, .., d$ define $\theta^{(1)}_i, \theta^{(2)}_i$ for each option of labels:

i. $y_i = 1, y'_i = 1$ - define $\theta^{(1)}_i = -2, \theta^{(2)}_i = 2$

ii. $y_i = -1, y'_i = 1$ - define $\theta^{(1)}_i = -0.5, \theta^{(2)}_i = 2$

iii. $y_i = 1, y'_i = -1$ - define $\theta^{(1)}_i = -2, \theta^{(2)}_i = 0.5$

iv. $y_i = -1, y'_i = -1$ - define $\theta^{(1)}_i = -0.5, \theta^{(2)}_i = 0.5$

Notice that h_θ obtaind this way is in $\mathcal{H}^d_{rect}$ since its d - dimention rectangle. In addition, notice that the ith sample is in this rectangle iff the ith component of the sample is in $\left(\theta^{(1)}_i, \theta^{(2)}_i\right)$. Now, 0 always in this range, and since all the components of sample i are 0 beside the ith component, then from the construction above, we get correct labeling for all samples and so we successfully shattered S with $\mathcal{H}^d_{rect}$.

4. Prove that $VCdim\left(\mathcal{H}_{rect}^{d}\right) = 2d$.

   Solution - Using section (b) it's suffice to prove $VCdim\left(\mathcal{H}_{rect}^{d}\right) < 2d + 1$. Let S be any sample set with $2d + 1$ points. Define the rectangle with $\theta_i^{(1)} = \min_i$ and $\theta_i^{(2)} = \max_i$ where $\min_i$ is the minimum value of all ith components of the samples and $\max_i$ the maximum value of all ith components of the samples. Now, notice that from the pigeonhole principle, since we have $2d+1$ points, at least one point is right inside this rectangle. If we lable this point by -1 and all the rest with $+1$ we clearly can't find a rectangle that separates this labeling correctly. Thus, the by defenition $VCdim\left(\mathcal{H}_{rect}^{d}\right) < 2d + 1$ as required.