Introduction to Machine Learning Course

Dry 3 – SVM & PAC learning

Submitted individually by Thursday, 20.05, at 23:59. Each day of delay costs 5 points.

You may answer in Hebrew or English and write on a computer or by hand (but be clear).

Please submit a PDF file named like your ID number, e.g., 123456789.pdf.

Bonus opportunities (maximal grade is 100):

a. Writing on a computer (using LyX/LaTeX, Word + Equation tool, etc.) – adds 3 pts.

b. Writing in English – adds 2 pts.

1. Recall the optimization formulations for SVM:

Hard SVM

$$\operatorname*{argmin}_{w\in\mathbb{R}^d} \|w\|_2^2 \quad \text{s.t.} \quad y_i \cdot w^\top x_i \geq 1, \quad \forall i \in [m]$$

Soft SVM

$$\operatorname*{argmin}_{w\in\mathbb{R}^d} \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i \cdot w^\top x_i\} + \lambda\|w\|_2^2$$

For the (homogeneous) linearly separable case:

1.1. When $\lambda \to \infty$, to which solution will the soft SVM converge?

1.2. 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.

2. Let $K_1(u,v) = \langle \phi_1(u), \phi_1(v) \rangle,\ K_2(u,v) = \langle \phi_2(u), \phi_2(v) \rangle$ be two kernels with corresponding feature mappings $\phi_1: \mathcal{X} \to \mathbb{R}^{n_1}, \phi_2: \mathcal{X} \to \mathbb{R}^{n_2}$ where $n_1, n_2 \in \mathbb{N}$.

Notice that K_1, K_2 are valid (i.e., well-defined) kernels since each of them can be written as an inner product of some mapping of u and v.

Prove that $K_3(u,v) = K_1(u,v) + K_2(u,v)$ is a valid kernel. You should do so by showing a feature mapping $\phi_3: \mathcal{X} \to \mathbb{R}^{n_3}$ for some $n_3 \in \mathbb{N}$, such that $K_3(u,v) = \langle \phi_3(u), \phi_3(v) \rangle$.

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

$$\mathcal{X} = \mathbb{R}^d,\ \ \mathcal{H}_{\text{rect}}^d = \left\{ h_\theta \,\middle|\, \forall i \in [d]: \theta_i^{(1)} < \theta_i^{(2)} \right\}$$

where $\theta^{(1)}, \theta^{(2)} \in \mathbb{R}^d, \theta = \left(\theta^{(1)}, \theta^{(2)}\right)$ and $h_\theta(\boldsymbol{x}) = \begin{cases} +1, & \Lambda_{i\in[d]}\left(\theta_i^{(1)} \leq x_i \leq \theta_i^{(2)}\right) \\ -1, & \text{otherwise} \end{cases}$.

We saw on tutorial 05 that $\text{VCdim}(\mathcal{H}_{\text{rect}}^2) = 4$.

For the general d-dimensional case:

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

3.2. Prove that $\text{VCdim}\left(\mathcal{H}_{\text{rect}}^d\right) \geq 2d$

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