Introduction to Machine Learning Course

Short HW3 – SVM, Optimization, and PAC learning

Submitted individually by Sunday, 18.12, at 23:59.

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 (maximal grade is 100): Writing on a computer (using LyX/LaTeX, Word + Equation tool, etc.) = 2 pts.

1. **VC-dimension exercises:**

   1.1. In tutorial 05, we defined the hypothesis class of axis aligned rectangles (or cuboids) in $\mathbb{R}^2$.

   $$\mathcal{X}=\mathbb{R}^2,\ \ \mathcal{H}_{\text{rect}}=\{h_\theta \mid \theta=(a_1,a_2,b_1,b_2)\in\mathbb{R}^4,\quad a_1<a_2,\ \ b_1<b_2\}$$

   where a single hypothesis is defined by $h_\theta(\boldsymbol{x})=\begin{cases}+1, & (a_1\le x_1\le a_2)\wedge(b_1\le x_2\le b_2)\\ -1, & \text{otherwise}\end{cases}$.

   We saw that $\mathrm{VCdim}(\mathcal{H}_{\text{rect}})\ge 4$.

   Rigorously prove that $\mathrm{VCdim}(\mathcal{H}_{\text{rect}})<5$ (thus proving that $\mathrm{VCdim}(\mathcal{H}_{\text{rect}})=4$).

   1.2. Prove that the VC-dimension is monotone:

   For any two hypothesis classes, if $\mathcal{H}_1\subseteq\mathcal{H}_2$ then $\mathrm{VCdim}(\mathcal{H}_1)\le\mathrm{VCdim}(\mathcal{H}_2)$.

   1.3. Using only the above, what can be said on the VC-dimension of $\mathcal{H}_{\text{DT}}$, the class of decision trees of at most depth 4 (recall slides 4-5 in Tutorial 03)?

   Prove your answer (to claim that $\mathcal{H}_1\subseteq\mathcal{H}_2$, you need to prove that $h\in\mathcal{H}_1\Rightarrow h\in\mathcal{H}_2$).

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 is a valid (i.e., well-defined) kernel since it can be written as an inner product of a mapping of u and v.

   Prove that $K_3(u,v)=4\cdot K_1(u,v)+9\cdot K_2(u,v)$ is a valid kernel. That is, propose a feature mapping $\phi_3:\mathcal{X}\to\mathbb{R}^{n_3}$ for some $n_3\in\mathbb{N}$, such that $K_3(u,v)=4\cdot K_1(u,v)+9\cdot K_2(u,v)=\langle\phi_3(u),\phi_3(v)\rangle$.

3. We will now prove that the following Soft-SVM problem is convex:

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

   Let $f,g:C\to\mathbb{R}$ be two convex functions defined over a convex set C.

   Lemma (no need to prove)**:** given a constant $\alpha\in\mathbb{R}_{\ge 0}$, the function $\alpha f(z)$ is convex w.r.t z.

   Lemma (no need to prove)**:** a sum of any number of convex functions is convex.

   3.1. Prove (by definition) that $q(z)\triangleq\max\{f(z),\ g(z)\}$ is convex w.r.t z.

   3.2. Using a rule from Tutorial 07, conclude that $\max\{0,1-y_i\boldsymbol{w}^\top\boldsymbol{x}_i\}$ is convex w.r.t $\boldsymbol{w}$.

   3.3. Using the above (and properties from Tutorial 07), conclude that the Soft-SVM optimization problem is convex w.r.t $\boldsymbol{w}$.