HW2

November 2, 2016

A. Rademacher Complexity

1.

$$
\begin{aligned}
\mathcal{R}'(h) &= \frac{1}{m}\mathbf{E}_{\sigma,S}\left[sup_{h\in H}|\sum_{i=1}^{m}\sigma_i h(x_i)|\right] \\
&\leq \frac{1}{m}\mathbf{E}_{\sigma,S}\left[sup_{h\in H}\sum_{i=1}^{m}|\sigma_i||h(x_i)|\right] \\
&\leq \frac{1}{m}\mathbf{E}_S\left[m\ sup_{h\in H}|h(x)|\right] \\
&\leq \frac{1}{m}\sqrt{m\mathbf{E}_S\left[sup_{h\in H}(|h(x)|)^2\right]} \\
&= \frac{1}{m}\sqrt{m\mathbf{E}_{x\sim D}\left[h^2(x)\right]} \\
&= \sqrt{\frac{\mathbf{E}_{x\sim D}\left[h^2(x)\right]}{m}}
\end{aligned}
$$

2. •

$$
\begin{aligned}
\mathcal{R}_m(\alpha H) &= \mathbf{E}_\sigma\left[sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i\alpha h(x_i)\right] \\
&= \mathbf{E}_{\sigma'}\left[|\alpha|sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma'_i h(x_i)\right] \\
&= |\alpha|\mathcal{R}_m(H)
\end{aligned}
$$

•

$$
\begin{aligned}
\mathcal{R}_m(H+H') &= \mathbf{E}_\sigma\left[sup_{h\in H,h'\in H'}\frac{1}{m}\sum_{i=1}^{m}\sigma_i(h(x_i)+h'(x))\right] \\
(subadditivity) &\leq \mathbf{E}_\sigma\left[sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)+sup_{h'\in H'}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h'(x)\right] \\
&= \mathcal{R}_m(H)+\mathcal{R}_m(H')
\end{aligned}
$$

-

$$
\begin{aligned}
\mathcal{R}_m(\{max(h,h'): h \in H, h' \in H'\}) &= \mathbf{E}_\sigma \left[sup_{h\in H, h' \in H'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i max((h(x_i), h'(x)) \right] \\
&= \mathbf{E}_\sigma \left[sup_{h\in H, h' \in H'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \frac{1}{2} [h(x_i) + h'(x_i) + |h(x_i) - h'(x_i)|] \right] \\
\text{(Lipschitz)} &\leq \frac{1}{2} \mathbf{E}_\sigma \left[sup_{h\in H, h' \in H'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i [h(x_i) + h'(x_i) + |h(x_i) - h'(x_i)|] \right] \\
\text{(subadditivity)} &\leq \frac{1}{2} \mathbf{E}_\sigma \Big[sup_{h\in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(x_i) + sup_{h' \in H'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h'(x_i) \\
&+ sup_{h\in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(x_i) + sup_{h' \in H'} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h'(x_i) \Big] \\
&= \mathcal{R}_m(H) + \mathcal{R}_m(H')
\end{aligned}
$$

B. VC-dimension

1. Unlike *intervals on real line* in lecture, any three points can be shattered by the subset of real line. However, in four-point case, (+,-,+,-) cannot be shattered. So the VC-dimension is 3.

2. (a) Counter example: (+,-,-,-)

$$
\begin{cases}
\omega x = (2k\pi, 2k\pi + \pi) \\
2\omega x = (2k\pi + \pi, 2k\pi + 2\pi) \\
3\omega x = (2k\pi + \pi, 2k\pi + 2\pi) \\
4\omega x = (2k\pi + \pi, 2k\pi + 2\pi)
\end{cases}
\Rightarrow
\begin{array}{l}
\omega x = (2k\pi, 2k\pi + \pi) \\
\omega x = (k\pi + \frac{1}{2}\pi, k\pi + \pi) \\
\omega x = (\frac{2}{3}k\pi + \frac{1}{3}\pi, \frac{2}{3}k\pi + \frac{2}{3}\pi) \\
\omega x = (\frac{1}{2}k\pi + \frac{1}{4}\pi, \frac{1}{2}k\pi + \frac{1}{2}\pi)
\end{array}
\tag{1}
$$

   Note, there is no intersection among these four regions no matter how k is chosen. So for any $x \in \mathbb{R}$, $x, 2x, 3x, 4x$ cannot be shattered by the family of sine functions.

   (b) Similarly,

$$\omega \in 2^m(2k\pi, 2k\pi + \pi)$$

   for positive points, and

$$\omega \in 2^m(2k\pi + \pi, 2k\pi + 2\pi)$$

   for negative ones.

   For $m = 0$, for a particular point, ω could ony be in half of the space, while for $m > 0$, positive or negative, ω covers all the space of the value of sine functions. That is to say, $\forall \omega \in \mathbb{N}$, there is a selection of ω such that $\{2^{-m}\}$ can be fully shattered.

C. Support Vector Machine

1. The splitting, scaling, training and testing operations are written as script.

2. In my experient, $C^* = 4$

   I know my script goes wrong because d value does not matter. I tested on cross validation and test data, and I also checked the number of SV, they were all the same, with value 15, 15, 1919.

3. If the penalty because $||\xi||_\infty$, the only difference is $\nabla_{\xi_i} L$ in KKT, which becomes $\alpha_i + \beta_i = C\mathbf{1}_{\xi_i = max_j\ \xi_j}$. The goal function of the dual problem does not change, only contraint $0 \leq \alpha_i \leq C$ changes to $0 \leq ||\alpha||_\infty \leq C$.

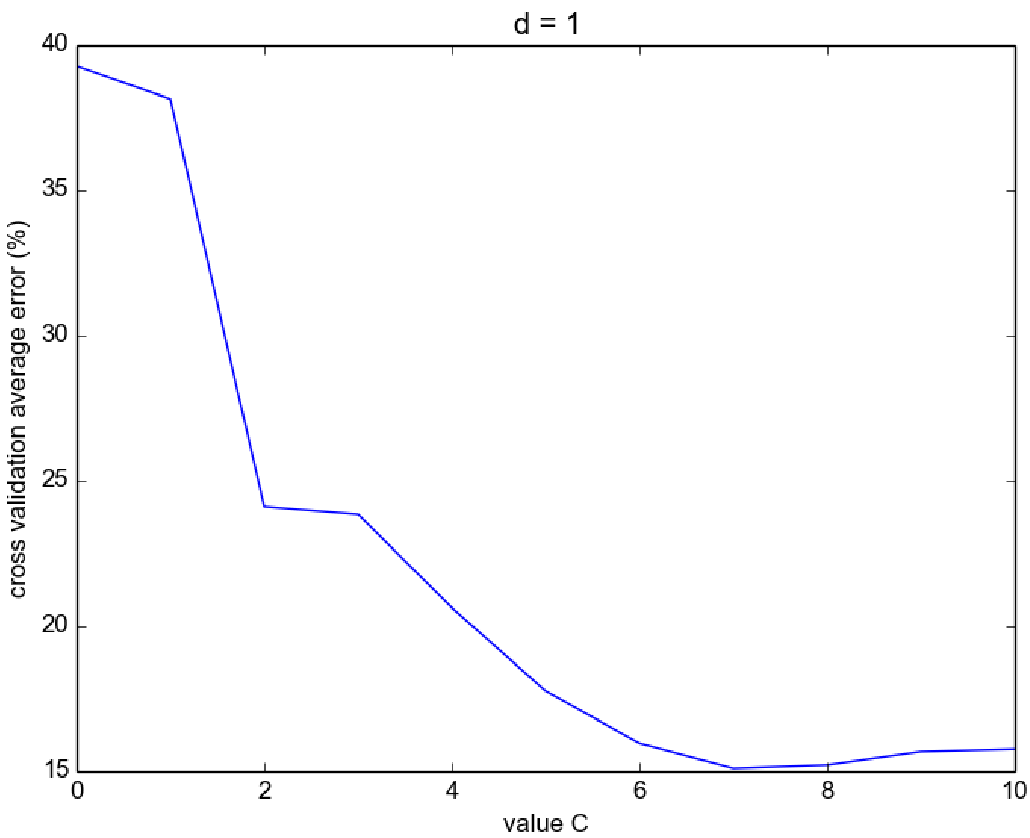

Figure 1: Average cross validation error versus C.