Recall that the solution of SVM is $f(x) = \sum_{i=1}^n \alpha_i K(x_i, x)$, where $(x_1, y_1), \ldots, (x_n, y_n)$ – data, with $y_i \in \{-1, 1\}$. The label is predicted by $\text{sign}(f(x))$ and $\mathbb{P}(yf(x) \le 0)$ is *misclassification error*.

Let $\mathcal{H} = \mathcal{H}((x_1, y_1), \ldots, (x_n, y_n))$ be random collection of functions, with card $\mathcal{H} \le \mathcal{N}(n)$. Also, assume that for any $h \in \mathcal{H}$, $-h \in \mathcal{H}$ so that α can be positive.

Define

$$\mathcal{F} = \left\{ \sum_{i=1}^T \lambda_i h_i,\ T \ge 1,\ \lambda_i \ge 0,\ \sum_{i=1}^T \lambda_i = 1,\ h_i \in \mathcal{H} \right\}.$$

For SVM, $\mathcal{H} = \{\pm K(x_i, x) : i = 1, \ldots, n\}$ and card $\mathcal{H} \le 2n$.

Recall margin-sparsity bound (voting classifiers): algorithm outputs $f = \sum_{i=1}^T \lambda_i h_i$. Take random approximation $g(x) = \frac{1}{k}\sum_{j=1}^k Y_j(x)$, where $Y_1, \ldots, Y_k$ i.i.d with $\mathbb{P}(Y_j = h_i) = \lambda_i$, $\mathbb{E}Y_j(x) = f(x)$.

Fix $\delta > 0$.

$$\begin{aligned}
\mathbb{P}(yf(x) \le 0) &= \mathbb{P}(yf(x) \le 0, yg(x) \le \delta) + \mathbb{P}(yf(x) \le 0, yg(x) > \delta) \\
&\le \mathbb{P}(yg(x) \le \delta) + \mathbb{E}_{x,y}\mathbb{P}_Y\left(y\frac{1}{k}\sum_{j=1}^k Y_j(x) > \delta,\ y\mathbb{E}_Y Y_1(x) \le 0 \right) \\
&\le \mathbb{P}(yg(x) \le \delta) + \mathbb{E}_{x,y}\mathbb{P}_Y\left(\frac{1}{k}\sum_{j=1}^k (yY_j(x) - \mathbb{E}(yY_j(x))) \ge \delta \right) \\
&\le \text{(by Hoeffding) } \mathbb{P}(yg(x) \le \delta) + \mathbb{E}_{x,y} e^{-k\delta^2/2} \\
&= \mathbb{P}(yg(x) \le \delta) + e^{-k\delta^2/2} \\
&= \mathbb{E}_Y \mathbb{P}_{x,y}(yg(x) \le \delta) + e^{-k\delta^2/2}
\end{aligned}$$

Similarly to what we did before, on the data

$$\mathbb{E}_Y\left[\frac{1}{n}\sum_{i=1}^n I(y_i g(x_i) \le \delta) \right] \le \frac{1}{n}\sum_{i=1}^n I(y_i f(x_i) \le 2\delta) + e^{-k\delta^2/2}$$

Can we bound

$$\mathbb{P}_{x,y}(yg(x) \le \delta) - \frac{1}{n}\sum_{i=1}^n I(y_i g(x_i) \le \delta)$$

for any g?

Define

$$\mathcal{C} = \{\{yg(x) \le \delta\},\ g \in \mathcal{F}_k,\ \delta \in [-1, 1]\}$$

where

$$\mathcal{F}_k = \left\{ \frac{1}{k}\sum_{j=1}^k h_j(x) :\ h_j \in \mathcal{H} \right\}$$

Note that $\mathcal{H}(x_1, \ldots, x_n) \subseteq \mathcal{H}(x_1, \ldots, x_n, x_{n+1})$ and $\mathcal{H}(\pi(x_1, \ldots, x_n)) = \mathcal{H}(x_1, \ldots, x_n)$.

In the last lecture, we proved

$$\mathbb{P}_{x,y}\left(\sup_{C\in\mathcal{C}} \frac{\mathbb{P}\left(C\right) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C)}{\sqrt{\mathbb{P}\left(C\right)}} \geq t\right) \leq 4G(2n)e^{-\frac{nt^2}{2}}$$

where

$$G(n) = \mathbb{E}\triangle_{\mathcal{C}(x_1,\ldots,x_n)}(x_1,\ldots,x_n).$$

How many different g's are there? At most card $\mathcal{F}_k \leq \mathcal{N}(n)^k$. For a fixed g,

$$\text{card}\ \{\{yg(x) \leq \delta\} \cap \{x_1,\ldots,x_n\},\ \delta \in [-1,1]\} \leq (n+1).$$

Indeed, we can order $y_1 g(x_1),\ldots,y_n g(x_n) \to y_{i_1} g(x_{i_1}) \leq \ldots \leq y_{i_n} g(x_{i_n})$ and level δ can be anywhere along this chain.

Hence,

$$\triangle_{\mathcal{C}(x_1,\ldots,x_n)}(x_1,\ldots,x_n) \leq \mathcal{N}(n)^k(n+1).$$

$$\begin{aligned}\mathbb{P}_{x,y}\left(\sup_{C\in\mathcal{C}} \frac{\mathbb{P}\left(C\right) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C)}{\sqrt{\mathbb{P}\left(C\right)}} \geq t\right) &\leq 4G(2n)e^{-\frac{nt^2}{2}} \\ &\leq 4\mathcal{N}(2n)^k(2n+1)e^{-\frac{nt^2}{2}}\end{aligned}$$

Setting the above bound to e^{-u} and solving for t, we get

$$t = \sqrt{\frac{2}{n}(u + k\log\mathcal{N}(2n) + \log(8n+4))}$$

So, with probability at least $1 - e^{-u}$, for all C

$$\frac{\left(\mathbb{P}\left(C\right) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C)\right)^2}{\mathbb{P}\left(C\right)} \leq \frac{2}{n}\left(u + k\log\mathcal{N}(2n) + \log(8n+4)\right).$$

In particular,

$$\frac{\left(\mathbb{P}\left(yg(x) \leq \delta\right) - \frac{1}{n}\sum_{i=1}^n I(y_i g(x_i) \leq \delta)\right)^2}{\mathbb{P}\left(yg(x) \leq \delta\right)} \leq \frac{2}{n}\left(u + k\log\mathcal{N}(2n) + \log(8n+4)\right).$$

Since $\frac{(x-y)^2}{x}$ is convex with respect to (x,y),

$$\begin{aligned}&\frac{\left(\mathbb{E}_Y\mathbb{P}_{x,y}\left(yg(x) \leq \delta\right) - \mathbb{E}_Y\frac{1}{n}\sum_{i=1}^n I(y_i g(x_i) \leq \delta)\right)^2}{\mathbb{E}_Y\mathbb{P}_{x,y}\left(yg(x) \leq \delta\right)} \\ &\leq \mathbb{E}_Y\frac{\left(\mathbb{P}\left(yg(x) \leq \delta\right) - \frac{1}{n}\sum_{i=1}^n I(y_i g(x_i) \leq \delta)\right)^2}{\mathbb{P}\left(yg(x) \leq \delta\right)} \qquad (32.1) \\ &\leq \frac{2}{n}\left(u + k\log\mathcal{N}(2n) + \log(8n+4)\right).\end{aligned}$$

Recall that

$$\mathbb{P}\left(yf(x) \leq 0\right) \leq \mathbb{E}_Y\mathbb{P}\left(yg(x) \leq \delta\right) + e^{-k\delta^2/2} \qquad (32.2)$$

and

$$\mathbb{E}_Y \frac{1}{n}\sum_{i=1}^{n} I(y_i g(x_i) \le \delta) \le \frac{1}{n}\sum_{i=1}^{n} I(y_i f(x_i) \le 2\delta) + e^{-k\delta^2/2}. \tag{32.3}$$

Choose k such that $e^{-k\delta^2/2} = \frac{1}{n}$, i.e. $k = \frac{2\log n}{\delta^2}$. Plug (32.2) and (32.3) into (32.1) (look at $\frac{(a-b)^2}{a}$). Hence,

$$\frac{\left(\mathbb{P}\left(yf(x) \le 0\right) - \frac{2}{n} - \frac{1}{n}\sum_{i=1}^{n} I(y_i f(x_i) \le 2\delta)\right)^2}{\mathbb{P}\left(yf(x) \le 0\right) - \frac{2}{n}} \le \frac{2}{n}\left(u + \frac{2\log n}{\delta^2}\log \mathcal{N}(2n) + \log(8n+4)\right)$$

with probability at least $1 - e^{-u}$.

Recall that for SVM, $\mathcal{N}(n) = \text{card}\,\{\pm K(x_i, x)\} \le 2n$.