As in the previous lecture, let $\mathcal{F} = \{(w, \phi(x))_{\mathcal{H}}, \|w\| \le 1\}$, where $\phi(x) = (\sqrt{\lambda_i}\phi_i(x))_{i\ge 1}$, $\mathcal{X} \subset \mathbb{R}^d$.
Define $d(f,g) = \|f-g\|_\infty = \sup_{x\in\mathcal{X}} |f(x) - g(x)|$.
The following theorem appears in Cucker & Smale:

Theorem 31.1. *$\forall h \ge d$,*

$$\log \mathcal{N}(\mathcal{F}, \varepsilon, d) \le \left(\frac{C_h}{\varepsilon}\right)^{\frac{2d}{h}}$$

where C_h is a constant.

Note that for any $x_1, \ldots, x_n$,

$$d_x(f,g) = \left(\frac{1}{n}\sum_{i=1}^n (f(x_i) - g(x_i))^2\right)^{1/2} \le d(f,g) = \sup_x |f(x) - g(x)| \le \varepsilon.$$

Hence,

$$\mathcal{N}(\mathcal{F}, \varepsilon, d_x) \le \mathcal{N}(\mathcal{F}, \varepsilon, d).$$

Assume the loss function $\mathcal{L}(y, f(x)) = (y - f(x))^2$. The *loss class* is defined as

$$\mathcal{L}(y, F) = \{(y - f(x))^2, f \in \mathcal{F}\}.$$

Suppose $|y - f(x)| \le M$. Then

$$|(y - f(x))^2 - (y - g(x))^2| \le 2M|f(x) - g(x)| \le \varepsilon.$$

So,

$$\mathcal{N}(\mathcal{L}(y,\mathcal{F}), \varepsilon, d_x) \le \mathcal{N}\left(\mathcal{F}, \frac{\varepsilon}{2M}, d_x\right)$$

and

$$\log \mathcal{N}(\mathcal{L}(y,\mathcal{F}), \varepsilon, d_x) \le \left(\frac{2MC_h}{\varepsilon}\right)^{\frac{2d}{h}} = \left(\frac{2MC_h}{\varepsilon}\right)^{\alpha}$$

$\alpha = \frac{2d}{h} < 2$ (see Homework 2, problem 4).
Now, we would like to use specific form of solution for SVM: $f(x) = \sum_{i=1}^n \alpha_i K(x_i, x)$, i.e. f belongs to a random subclass. We now prove a VC inequality for random collection of sets.
Let's consider $\mathcal{C}(x_1, \ldots, x_n) = \{C : C \subseteq \mathcal{X}\}$ - random collection of sets. Assume that $\mathcal{C}(x_1, \ldots, x_n)$ satisfies:

(1) $C(x_1, \ldots, x_n) \subseteq C(x_1, \ldots, x_n, x_{n+1})$
(2) $C(\pi(x_1, \ldots, x_n)) = C(x_1, \ldots, x_n)$ for any permutation π.

Let

$$\triangle_{\mathcal{C}}(x_1, \ldots, x_n) = \text{card}\ \{C \cap \{x_1, \ldots, x_n\}; C \in \mathcal{C}\}$$

and

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

Theorem 31.2.

$$\mathbb{P}\left(\sup_{C\in\mathcal{C}(x_1,\ldots,x_n)} \frac{\mathbb{P}(C) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C)}{\sqrt{\mathbb{P}(C)}} \geq t\right) \leq 4G(2n)e^{-\frac{nt^2}{4}}$$

Consider event

$$A_x = \left\{x = (x_1,\ldots,x_n) : \sup_{C\in\mathcal{C}(x_1,\ldots,x_n)} \frac{\mathbb{P}(C) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C)}{\sqrt{\mathbb{P}(C)}} \geq t\right\}$$

So, there exists $C_x \in \mathcal{C}(x_1,\ldots,x_n)$ such that

$$\frac{\mathbb{P}(C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\mathbb{P}(C_x)}} \geq t.$$

For $x'_1,\ldots,x'_n$, an independent copy of x,

$$\mathbb{P}_{x'}\left(\mathbb{P}(C_x) \leq \frac{1}{n}\sum_{i=1}^n I(x'_i\in C_x)\right) \geq \frac{1}{4}$$

if $\mathbb{P}(C_x) \geq \frac{1}{n}$ (which we can assume without loss of generality).

Together,

$$\mathbb{P}(C_x) \leq \frac{1}{n}\sum_{i=1}^n I(x'_i\in C_x)$$

and

$$\frac{\mathbb{P}(C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\mathbb{P}(C_x)}} \geq t$$

imply

$$\frac{\frac{1}{n}\sum_{i=1}^n I(x'_i\in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\frac{1}{2n}\sum_{i=1}^n (I(x'_i\in C_x) + I(x_i\in C_x))}} \geq t.$$

Indeed,

$$\begin{aligned}
0 < t &\leq \frac{\mathbb{P}(C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\mathbb{P}(C_x)}} \\
&\leq \frac{\mathbb{P}(C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\frac{1}{2}\left(\mathbb{P}(C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)\right)}} \\
&\leq \frac{\frac{1}{n}\sum_{i=1}^n I(x'_i\in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i\in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i\in C_x)\right)}}
\end{aligned}$$

Hence, multiplying by an indicator,

$$\begin{aligned}
\frac{1}{4} \cdot I(x \in A_x) &\le \mathbb{P}_{x'}\left(\mathbb{P}(C_x) \le \frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x)\right) \cdot I(x \in A_x) \\
&\le \mathbb{P}_{x'}\left(\frac{\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right) \\
&\le \mathbb{P}_{x'}\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n)} \frac{\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right)
\end{aligned}$$

Taking expectation with respect to x on both sides,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n)} \frac{\mathbb{P}(C) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C)}{\sqrt{\mathbb{P}(C)}} \ge t\right) \\
&\le 4\mathbb{P}\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n)} \frac{\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right) \\
&\le 4\mathbb{P}\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)} \frac{\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) - \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right) \\
&= 4\mathbb{P}\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)} \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i(I(x'_i \in C_x) - I(x_i \in C_x))}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right) \\
&= 4\mathbb{E}\mathbb{P}_\varepsilon\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)} \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i(I(x'_i \in C_x) - I(x_i \in C_x))}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right)
\end{aligned}$$

By Hoeffding,

$$\begin{aligned}
&4\mathbb{E}\mathbb{P}_\varepsilon\left(\sup_{C \in \mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)} \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i(I(x'_i \in C_x) - I(x_i \in C_x))}{\sqrt{\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n I(x'_i \in C_x) + \frac{1}{n}\sum_{i=1}^n I(x_i \in C_x)\right)}} \ge t\right) \\
&\le 4\mathbb{E}\triangle_{\mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)}(x_1,\ldots,x_n,x'_1,\ldots,x'_n) \cdot \exp\left(-\frac{t^2}{2\sum\left(\frac{\frac{1}{n}(I(x'_i \in C_x) - I(x_i \in C_x))}{\sqrt{\frac{1}{2n}\sum_{i=1}^n(I(x'_i \in C_x) + I(x_i \in C_x))}}\right)^2}\right) \\
&\le 4\mathbb{E}\triangle_{\mathcal{C}(x_1,\ldots,x_n,x'_1,\ldots,x'_n)}(x_1,\ldots,x_n,x'_1,\ldots,x'_n) \cdot e^{-\frac{nt^2}{4}} \\
&= 4G(2n)e^{-\frac{nt^2}{4}}
\end{aligned}$$