23.1 VC Subgraph Classes

For binary classifiers, we have the following bound on the excess risk

$$\mathcal{E}(\hat{g}_n) \le 2 \sup_{g \in G} |P_n g - Pg| + \mathcal{E}(\overline{g}),$$

where

$$\widehat{g}_n = \operatorname*{argmin}_{g \in G} P_n g = \operatorname*{argmin}_{g \in G} \frac{1}{n} \sum_{j=1}^{n} I\{Y_j \neq g(X_j)\}$$

and $\overline{g}$ is the Bayes classifier.

Problem 1. Let $\mathcal{F} = \{f : S \to \mathbb{R}\}$ be a class of functions. What is the size of

$$\mathbb{E}\, ||P_n - P||_{\mathcal{F}} = \mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{j=1}^{n} f(X_j) - \mathbb{E} f(x) \right| ?$$

Given $f : S \to \mathbb{R}$, let Γ_f be its "subgraph", i.e.

$$\Gamma_f = \{(x, t) \in S \times \mathbb{R} : \, |f(x)| \ge t\}.$$

Definition 1. The class $\mathcal{F}$ is called VC-subgraph if the collection of subgraphs of $f \in \mathcal{F}$ is a VC class.

Example 1. Let $P_{k,d}$ be a class of polynomials of d variables of degree at most k, $\mathcal{C}$ a VC class.

$$\mathcal{F} := \left\{ \sum_{j=1}^{m} p_j(x) I_{C_j} \, : \, p_j \in P_{k,d}, \, C_j \in \mathcal{C} \right\}.$$

Piecewise polynomials on the C_j.

Example 2. $S = \mathbb{R}^d$, φ a monotone function. Let

$$\mathcal{F} = \{\phi(|| \cdot - \theta||_2) : \, \theta \in \mathbb{R}^d\}.$$

This is a useful class for kernel density estimators; since their variation is bounded, we can write them as a difference of two monotone functions. We can regard the process generated by shifts of the form above. *Proof:* Consider the sets of the form $\{(x,t) : 0 \le \varphi(||x-\theta||_2) \le t\}$, i.e. the top parts of the subgraphs. Then

$$\begin{aligned} 0 \le t \le \varphi(||x-\theta||_2) &\Leftrightarrow 0 \le \varphi^{-1}(t) \le ||x-\theta||_2 \\ &\Leftrightarrow \varphi^{-1}(t)^2 \le ||x-\theta||_2^2 \\ &\Leftrightarrow \sum_{j=1}^{d}(x_j-\theta_j)^2 - \varphi^{-1}(t)^2 \ge 0 \end{aligned}$$

which is the set of positivity of a function $f(x_1, \ldots, x_d)$ which is a polynomial of degree 2. Hence, the VC dimension of $\Gamma_{\mathcal{F}}$ has order d^2.

Goal: bound $\mathbb{E}||P_n - P||_{\mathcal{F}}$.

Note that

$$\mathbb{E}||P_n - P||_{\mathcal{F}} \le 2\mathbb{E}_X\mathbb{E}_\varepsilon||R_n||_{\mathcal{F}} = 2\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum \varepsilon_j f(X_j)\right| \le 2\cdot 12\sqrt{2}\mathbb{E}_X\int_0^\infty \sqrt{\log N(\mathcal{F}, d_n, \varepsilon)}\, d\varepsilon,$$

where $d_n^2(f_1, f_2) = \frac{1}{n}\sum_{j=1}^n (f_1(X_j) - f_2(X_j))^2$.

Problem 2. How to estimate $N(\mathcal{F}, d_n, \varepsilon)$? We will bound it via $\sup_Q N(\mathcal{F}, L_2(Q), \varepsilon)$.

Given a class $\mathcal{F}$, define $F(x) = \sup_{f\in\mathcal{F}}|f(x)|$, the envelope of $\mathcal{F}$. Consider the space $S \times \mathbb{R}$ equipped with measure $Q \times \Lambda$, where Λ is the Lebesgue measure. Note that for $f, g \in \mathcal{F}$,

$$\begin{aligned} ||f-g||_{L_1(Q)} &= \int_S |f-g|\, dQ \\ &= \int_S \int_0^{|f-g|} 1\, dt\, dQ \\ &= \int_S \int_0^\infty I\{0 \le t \le |f-g|\}\, dt\, dQ \\ &= \int_0^\infty \int_S I\{0 \le t \le |f-g|\}\, dQ\, dt \\ &= (Q \times \Lambda)(\Gamma_f \Delta \Gamma_g). \end{aligned}$$

We normalize it by dividing by $2\int F\, dQ$. We name the result $\mu(\Gamma_f\Delta\Gamma_g)$. μ is a probability measure on $\mathcal{F}$; $\mu(\Gamma_F\Delta\Gamma_F) = 1$.