From Rademacher Complexity to Agnostic PAC Learning

- Denote examples as $z = (x, y)$, and assume a loss function $\ell(h, z)$.
- Our goal is to find $h$ with small $L_D(h) = E_{z \in D}[\ell(h, z)]$.
- Define $F = \{f_h \mid f_h(z) = \ell(h, z)\}$.
- Given $f \in F$, a sample $S$ induces a vector of losses $f(S) = (f(z_1), \ldots, f(z_m))$.
- Define $G(S) = \{f(S) \mid f \in F\}$.

McDiarmid’s Inequality: If $f(x_1, \ldots, x_m)$ changes by at most $c_i$ when we replace the $i$th input with an arbitrary element, and $x_1, \ldots, x_m$ are independent then

$$p(f(\ldots) - E[f(\ldots)] > \epsilon) \leq e^{-2\epsilon^2/\sum c_i^2}$$

Theorem: assuming $|\ell(h, z)| \leq c$, with prob $> 1 - \delta$,

$$L_D(ERM(S)) - L_D(h^*) \leq 2R(G(S)) + 4\sqrt{\frac{2\epsilon^2 \ln(4/\delta)}{m}}$$

Massart Lemma: for finite $A$,

$$R(A) \leq \max_{a \in A} \|a - \bar{a}\| \frac{1}{m} \sqrt{2 \ln |A|}$$

Theorem: Let $VCD(C) = d$. Let $S$ be an IID sample of size

$$m \geq \frac{8}{\epsilon^2}[8 \ln(\Phi_d(2m)) + 32 \ln(4/\delta)]$$

Then, with probability at least $1 - \delta$,

$$L_D(ERM(S)) - L_D(h^*) \leq \epsilon$$