EN.601.774 Theory of Replicable ML Spring 2025

Lecture 2

Instructor: *Jess Sorrell* Scribe: *Jess Sorrell*

Last time we considered the problem of "replicating" an estimate of the 0-1 loss of a given model h on distribution D. We wanted to show

$$\Pr_{S_1,S_2}[|\ell_{S_1}(h) - \ell_{S_2}(h)| \geq \varepsilon] \leq \delta,$$

and argued that it suffices to show

$$\Pr_{S_1,S_2}[|\ell_{S_1}(h) - \ell_D(h)| \geq \varepsilon/2] \leq \delta/2,$$

We proved Chebyshev's Inequality:

Theorem 0.1 (Chebyshev's Inequality)**.** *Let X be a random variable with non-zero variance $\sigma^2 = \mathsf{Var}(X)$. Then for any $\lambda > 0$*

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

And we applied it to the r.v. $\ell_{S_1}(h)$ to show that

$$\Pr_{S_1}[|\ell_{S_1}(h) - \ell_D(h)| \geq \varepsilon] \leq \frac{1}{4m\varepsilon^2}.$$

So if we want $\Pr_{S_1}[|\ell_{S_1}(h) - \ell_D(h)| \geq \varepsilon] < \delta$, we can take $m > \frac{1}{4\varepsilon^2\delta}$.

Great! So now we have some guarantee that, so long as we take our sample large enough (and so does the other team of researchers), replication efforts will be successful with good probability! Both research teams will end up with an empirical loss $\ell_S(h)$ that is close to its expectation $\ell_D(h)$, and therefore close to the other team's, except with probability 2δ. But we can do much, much better!

Theorem 0.2 (Hoeffding's Inequality)**.** *Let $X_1, X_2, \ldots, X_m$ be independent, bounded random variables with $X_i \in [a_i, b_i]$. Let $S_m = \sum_{i=1}^m X_i$. Then*

$$\Pr_{X_1,X_2,\ldots,X_m}[S_m \geq \mathbb{E}[S_m] + t] \leq e^{-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}}.$$

Note that this also applies to $S'_m = -\sum_{i=1}^m X_i$ and so

$$\Pr_{X_1,X_2,\ldots,X_m}[S'_m \geq \mathbb{E}[S'_m] + t] \leq e^{-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}}$$

$$\Pr_{X_1,X_2,\ldots,X_m}[S_m \leq \mathbb{E}[S_m] - t] \leq e^{-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}}$$

$$\Pr_{X_1,X_2,\ldots,X_m}[|S_m - \mathbb{E}[S_m]| \geq t] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2}}$$

Since the empirical 0-1 loss $\ell_{S_1}(h) = \frac{1}{m}\sum_{i=1}^{m} \ell(h(x_i), y_i))$, we have.

$$\Pr_{S\sim D^m}[|\ell_S(h) - \ell_D(h)| \geq t/m] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^{m}(b_i-a_i)^2}}$$

and so

$$\Pr_{S\sim D^m}[|\ell_S(h) - \ell_D(h)| \geq t] \leq 2e^{-2t^2 m}.$$

Then if we want to ensure $\Pr_{S\sim D^m}[|\ell_S(h) - \ell_D(h)| \geq \varepsilon] \leq \delta$, we can take

$$\begin{aligned} 2e^{-2\varepsilon^2 m} &\leq \ln(\delta) \\ \ln(2) - 2\varepsilon^2 m &\leq \ln(\delta) \\ 2\varepsilon^2 m &\geq -\ln(\delta/2) \\ m &\geq -\frac{\ln(\delta/2)}{2\varepsilon^2} \\ m &\geq \frac{\ln(2/\delta)}{2\varepsilon^2} \end{aligned}$$

many samples from D. This is only logarithmic in $1/\delta$, instead of linear!

We'll prove this theorem is 2 parts. We'll assume the following lemma (to be proved later).

Lemma 0.3 (Hoeffding's Lemma)**.** *Let X be a random variable such that $X \in [a, b]$. Then for any $\lambda > 0$,*

$$\mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}] \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

Proof. (Hoeffding's Inequality) From Markov's inequality, we know that for all $\lambda, t > 0$,

$$\begin{aligned} \Pr[S_m - \mathbb{E}[S_m] \geq t] &= \Pr[e^{\lambda(S_m-\mathbb{E}[S_m])} \geq e^{\lambda t}] \\ &\leq \frac{\mathbb{E}[e^{\lambda(S_m-\mathbb{E}[S_m])}]}{e^{\lambda t}} && \text{Markov's inequality} \\ &= \frac{\mathbb{E}[e^{\lambda(\sum_{i=1}^{m} X_i-\mathbb{E}[X_i])}]}{e^{\lambda t}} && \text{def of } S_m \text{ and linearity of } \mathbb{E} \\ &= \frac{\mathbb{E}[\prod_{i=1}^{m} e^{\lambda(X_i-\mathbb{E}[X_i])}]}{e^{\lambda t}} \\ &= \frac{\prod_{i=1}^{m} \mathbb{E}[e^{\lambda(X_i-\mathbb{E}[X_i])}]}{e^{\lambda t}} && \text{Independence of } X_i s \\ &\leq \frac{\prod_{i=1}^{m} e^{\frac{\lambda^2(b_i-a_i)^2}{8}}}{e^{\lambda t}} && \text{Hoeffding's lemma} \end{aligned}$$

We showed this is true for all $\lambda > 0$, so in particular is must be true for $\lambda = \frac{4t}{\sum_{i=1}^{m}(b_i-a_i)^2}$.

Then we have

$$\begin{aligned}\Pr[S_m - \mathbb{E}[S_m] \geq t] &\leq \frac{\prod_{i=1}^m e^{\frac{\lambda^2(b_i-a_i)^2}{8}}}{e^{\lambda t}} \\ &= \frac{e^{\frac{\lambda^2}{8}\sum_{i=1}^m (b_i-a_i)^2}}{e^{\lambda t}} \\ &= e^{\frac{\lambda t}{2}-\lambda t} \\ &= e^{-\frac{\lambda t}{2}} \\ &= e^{-\frac{2t^2}{\sum_{i=1}^m (b_i-a_i)^2}}\end{aligned}$$

□

Now it remains to prove the lemma.

Lemma 0.4 (Hoeffding's Lemma)**.** *Let X be a random variable such that $X \in [a,b]$. Then for any $\lambda > 0$,*

$$\mathbb{E}[e^{(\lambda X - \mathbb{E}[X])}] \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

Proof. We first define a new random variable $Z = X - \mathbb{E}[X]$ and note that $Z \in [c,d]$ for $l = a - \mathbb{E}[X], u = b - \mathbb{E}[X]$ (and so $b - a = u - l$).

By convexity of exp, we have that for all $z \in [c,d]$

$$e^{\lambda z} \leq \frac{u-z}{u-l}e^{\lambda l} + \frac{z-l}{u-c}e^{\lambda d}.$$

It follows that

$$\begin{aligned}\mathbb{E}[e^{\lambda Z}] &\leq \mathbb{E}\left[\frac{u-Z}{u-l}\right]e^{\lambda l} + \mathbb{E}\left[\frac{Z-l}{u-l}\right]e^{\lambda u} \\ &= \frac{ue^{\lambda l} - le^{\lambda u}}{u-l} && \mathbb{E}[Z] = 0.\end{aligned}$$

Where do we go now? If we could show that $\mathbb{E}[e^{\lambda Z}] \leq e^{F(\lambda(u-l))}$, for some function F, and then bound $F(x) \leq \frac{x^2}{8}$, we'd be set. So let's try to massage that last equality into the right form. We want to find F such that

$$e^{F(\lambda(u-l))} = \frac{ue^{\lambda l} - le^{\lambda u}}{u-l}$$

Note that $\lambda l = \frac{\lambda(u-l)l}{u-l}$ and $\lambda u = \frac{\lambda(u-l)u}{u-l}$. So writing $x = \lambda(u-l)$, our goal is to find F

such that

$$\begin{aligned}
e^{F(x)} &= \frac{ue^{\frac{xl}{u-l}} - le^{\frac{xu}{u-l}}}{u-l} \\
&= e^{\frac{xl}{u-l}}\left(\frac{u - le^{\frac{x(u-l)}{u-l}}}{u-l}\right) && \text{pull out } e^{\frac{xl}{u-l}} \\
&= e^{\frac{xl}{u-l}}\left(\frac{u-l+l-le^x}{u-l}\right) && \text{add } 0 \\
&= e^{\frac{xl}{u-l}}\left(1 + \frac{l-le^x}{u-l}\right)
\end{aligned}$$

So

$$\begin{aligned}
F(x) &= \ln\left(e^{\frac{xl}{u-l}}(1 + \tfrac{l-le^x}{u-l})\right) \\
&= \ln(e^{\frac{xl}{u-l}}) + \ln(1 + \tfrac{l-le^x}{u-l}) \\
&= \frac{xl}{u-l} + \ln(1 + \tfrac{l(1-e^x)}{u-l})
\end{aligned}$$

How do we show this is less than $\frac{\lambda^2(b-a)^2}{8}$? We'll apply Taylor's theorem to $F(x)$ around 0.

Theorem 0.5 (Taylor (specific to our applications))**.** *If a real-valued function F is twice-differentiable at $x = 0$, then there exists some $\gamma \in [0, 1]$ such that*

$$F(x) = F(0) + xF'(0) + \tfrac{x^2}{2}F''(\gamma x)$$

We have $F(0) = 0$. What about $F'(0)$?

$$\begin{aligned}
F'(x) &= \frac{l}{u-l} + \frac{\frac{df(x)}{dx}}{f(x)} && \text{for } f(x) = 1 + \tfrac{l(1-e^x)}{u-l} \\
\frac{df(x)}{dx} &= \frac{-le^x}{u-l} \\
\text{so } F'(0) &= \frac{l}{u-l} - \frac{\frac{l}{u-l}}{1} = 0.
\end{aligned}$$

One more!

$$\begin{aligned}
F''(x) &= \frac{d}{dx}\left(\frac{l}{u-l} + \frac{\frac{-le^x}{u-l}}{1+\frac{l(1-e^x)}{u-l}}\right) \\
&= \frac{d}{dx}\left(\frac{\frac{-le^x}{u-l}}{1+\frac{l(1-e^x)}{u-l}}\right) \\
&= \frac{d}{dx}\left(\frac{-le^x}{u-l+l(1-e^x)}\right) \\
&= \frac{d}{dx}\left(\frac{-le^x}{u-le^x}\right) \\
&= \frac{-lue^x}{(u-le^x)^2}
\end{aligned}$$

From the AMGM inequality, we know that $-lue^x \leq \frac{(u-le^x)^2}{4}$, so we have $F''(x) \leq 1/4$ for all x! Then Taylor's theorem tells us that there's some $\gamma \in [0,1]$ such that

$$\begin{aligned}
F(x) &= F(0) + xF'(0) + \tfrac{x^2}{2}F''(\gamma x) \\
&\leq \frac{x^2}{8}
\end{aligned}$$

and we're done!

Putting it all together we have

$$\mathbb{E}[e^{\lambda Z}] \leq e^{F(\lambda(u-l)} = e^{F(\lambda(b-c))} \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

□