Lecture 7: Bias, Var, Model Complexity

To precisely discuss evaluation need to keep track of what is random and what is fixed.

Consider training data $(\underset{\sim}{X}_n, Y_n) \overset{iid}{\sim} p$

Let $T = \{(\underset{\sim}{X}_n, Y_n)\}_{n=1}^{N}$ be training data

this is random

We use this training data to build a model

$$\hat{f} = \hat{f}_T$$

also random

Let $\underset{\sim}{X}_0, Y_0$ be indep (from training) sample from p

If M is some perf. metric (like L)

then let

T is fixed

$$\mathrm{Err}_T = \mathbb{E}[M(Y_0, \hat{f}(\underset{\sim}{X}_0)) \mid T]$$

what we'd really like to estimate

how close $\hat{f}(\underset{\sim}{X}_0)$ is to Y_0

= given specific T what is expected err. if we predict using $\hat{f}$ on independent sample $\underset{\sim}{X}_0, Y_0$

= gen. perf. using this training data T

This is quite difficult in practice to estimate.

Instead, typically easier to estimate

$$\mathrm{Err} = E_T[\mathrm{Err}_T]$$
$$= \mathbb{E}[M(Y_0, \hat{f}(\underset{\sim}{X}_0))]$$

= exp. gen. perf. over new samples AND over all possible training data

= exp. perf. of my model building process

Let $\overline{\mathrm{err}}$ = training error = $\frac{1}{N}\sum_n M(y_n, \hat{f}(\underset{\sim}{x}_n))$

Typically $\overline{\mathrm{err}} < \mathrm{Err}_T$ b/c we overfit

both calc. w/ fixed training

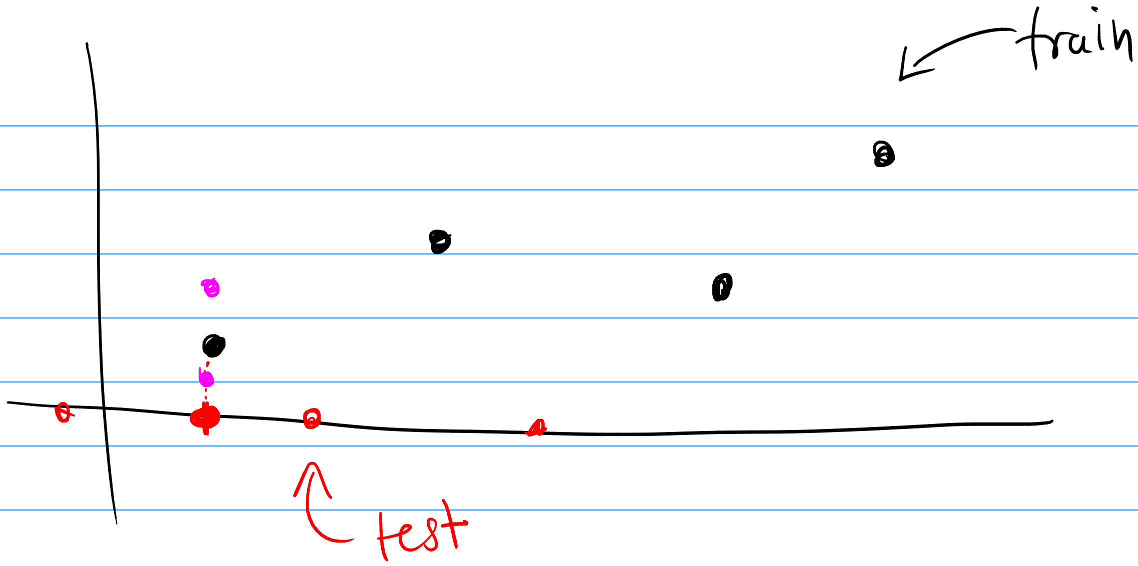
Parts of the problem

extra

truly

1. $\underset{\sim}{X}_0$ is random but $\underset{\sim}{x}_n$s are fixed
   So new $\underset{\sim}{X}_0$ may not be exactly the same as $\underset{\sim}{x}_n$s.

2. Y^0 is random so might not match Y_ns.



To simplify analysis, consider only Y^0 being random and define the in-sample error as

$$Err_{in} = \frac{1}{N}\sum_n \mathbb{E}_{Y^0}\left[M(Y_0, \hat{f}(\underset{\sim}{x}_n)) \,|\, T\right]$$

fixed at train pts

like Err_T but fixing $\underset{\sim}{x}_n$ at train pts

Let the optimism be

random (deps on Y^0)

$$op = Err_{in} - \overline{err}$$

in sample − training

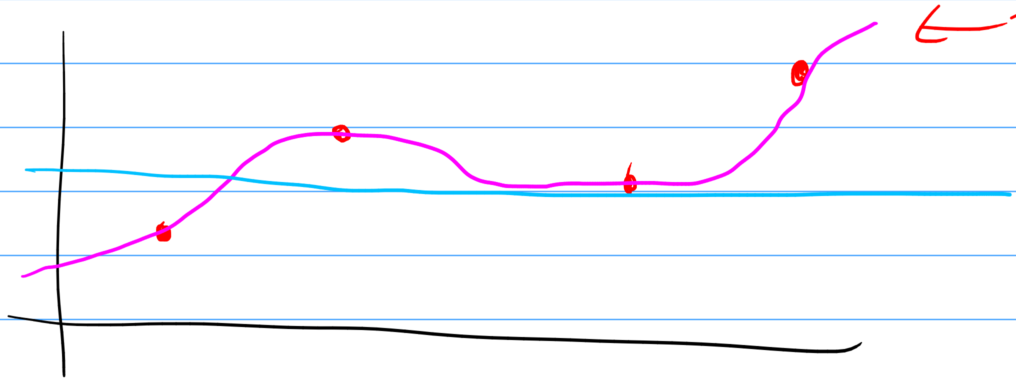
Typically $op > 0$ as $\overline{err}$ underestimates Err_{in} and if

$$\omega = \mathbb{E}[op]$$

= avg optimism

Generally,
$$\omega = \frac{2}{N} \sum_n \text{cov}(\hat{Y}_n, Y_n)$$

So ω is rel. to the avg. amt that Y_n affects $\hat{Y}_n$.



So
$$\mathbb{E}[\text{Err}_{in}] = \mathbb{E}[\overline{\text{err}}] + \omega$$

In some cases one can directly estmate ω as $\hat{\omega}$ in which case

$$\text{est. of Err}_{in} = \underbrace{\overline{\text{err}}}_{\text{know}} + \underbrace{\hat{\omega}}_{\text{estimate}}$$

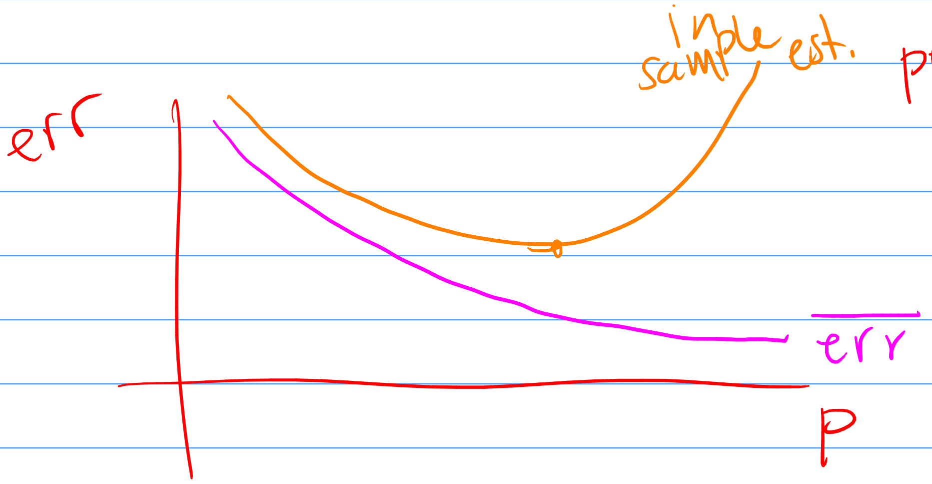
e.g. If we assume $Y = f(\underset{\sim}{X}) + \varepsilon$ when $\varepsilon \perp\!\!\!\perp \underset{\sim}{X}, Y$ and $\mathbb{E}\varepsilon = 0$, $Var(\varepsilon) = \sigma^2$

then $\omega = \frac{2 \overset{\#\text{ covs.}}{P}}{N} \sigma^2$

and so $\hat{\omega} = \frac{2P}{N} \hat{\sigma}^2$

est. of in-sample err $= \overline{err} + \frac{2P}{N}\hat{\sigma}^2$

penalty



Many ways to do this dep. on assumptions, C_p, AIC, BIC, adj. R^2

Hold on, this only estimates in-sample err — doesn't take into account randomness in $\underset{\sim}{X}_0$.

Yes, but often its good enough for model selection, b/c it still allows me to eval. rel. perf. among models.

If want to est. gen. perf. still need to do some kind of x-validation.

↳ estimating Err not Err_T