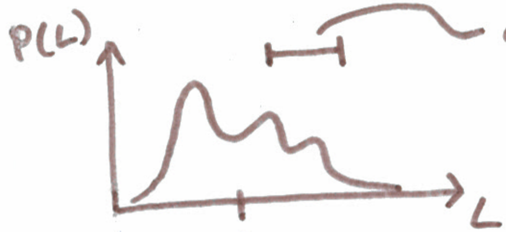
Standard error on mean test error

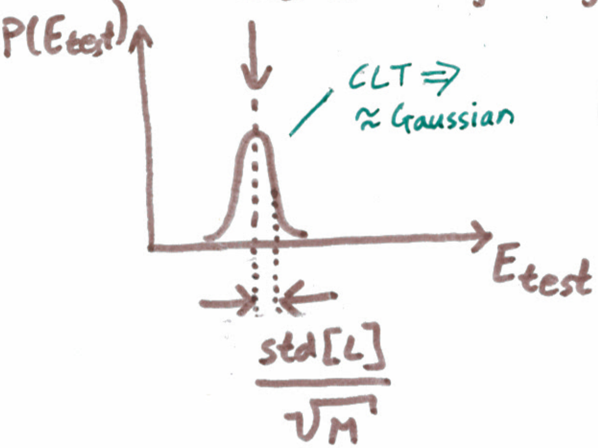


$\sigma_L = \mathrm{std}[L]$

(We don't know std[L], so estimate from data)

Loss on one random example

$\mathbb{E}[L] = E_{gen}$, generalization error



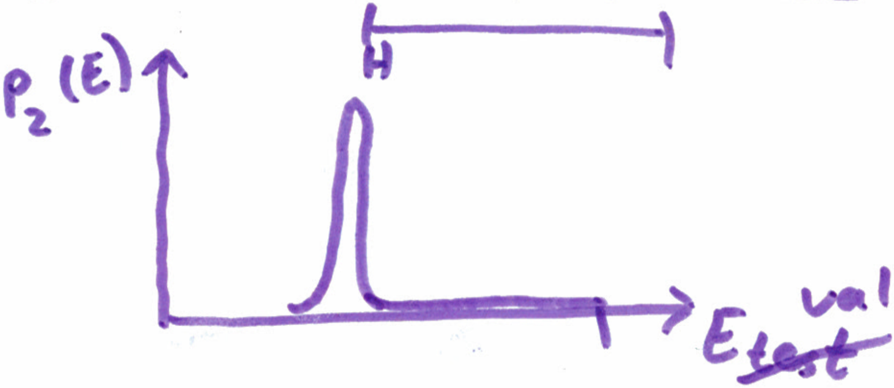
$$E_{test} = \frac{1}{M}\sum_m L_m$$

(Losses sampled on test set of size M)

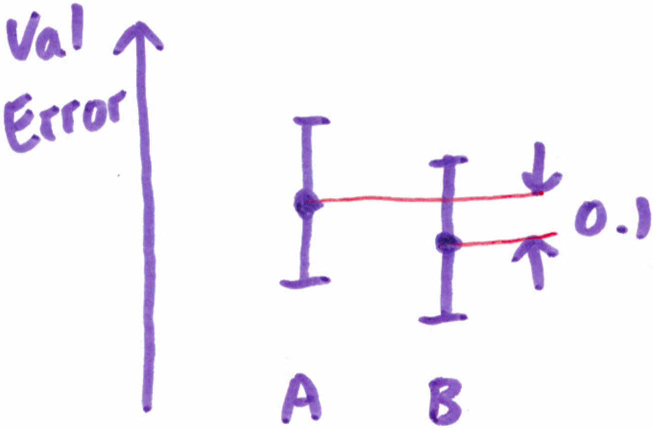
⇒ For a particular fitted model,

$$\text{"}E_{gen} = \underbrace{E_{test}}_{\text{A mean}} \pm \underbrace{\frac{\mathrm{std}[L]}{\sqrt{M}}}_{\text{"standard error on the mean"}}\text{"}$$

How variable is performance?



P_2: Distribution over $E_{\cancel{test}}^{val}$,
where I re-run the code.
(Or with different train sets)



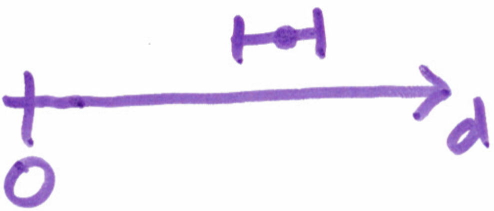
Q) Is ~~A~~ B better than ~~B~~ A?

<u>Paired Comparison</u>

Mean difference, $d = \frac{1}{M}\sum_m \left(L_m^{(A)} - L_m^{(B)}\right)$

Standard Error on mean:

$$= \frac{\mathrm{std}\left(L^{(A)} - L^{(B)}\right)}{\sqrt{M}}$$



Multivariate Gaussians

Sample $x_d \sim \mathcal{N}(0,1)$ indep. $d=1\ldots D$

$$p(\underline{x}) = \prod_d p(x_d) \text{ because indep.}$$

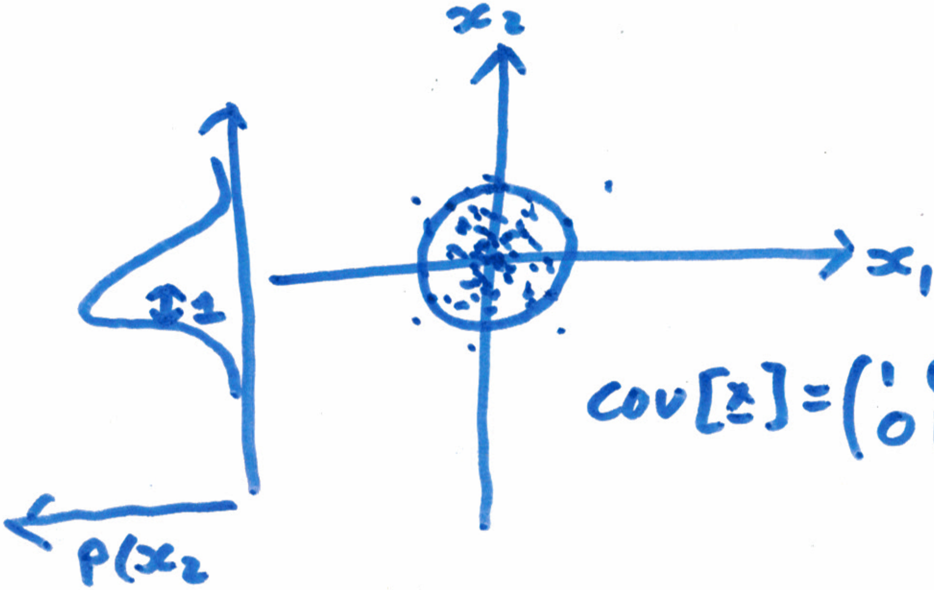
$$= \prod_d \mathcal{N}(x_d; 0, 1)$$

$$= \prod_d \frac{1}{\sqrt{2\pi}} e^{-x_d^2/2}$$

Sum
Not Σ covarian

$$= \frac{1}{(2\pi)^{D/2}} e^{-\sum_{d=1}^{D} x_d^2/2}$$

$$= \frac{1}{(2\pi)^{D/2}} e^{-\underline{x}^\top \underline{x}/2}$$



$$\mathrm{cov}[\underline{x}] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

e.g.

$$\mathrm{cov}[\underline{x}] = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

COVARIANCE

$\text{cov}[\underline{x}]$ is a $D \times D$ matix

$$\text{cov}[\underline{x}]_{ij} = \mathbb{E}[x_i x_j] - \mathbb{E}[x_i]\,\mathbb{E}[x_j]$$

$$\text{cov}[\underline{x}] = \mathbb{E}[\underset{D\times 1}{\underline{x}}\,\underset{1\times D}{\underline{x}^\top}] - \mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]^\top$$