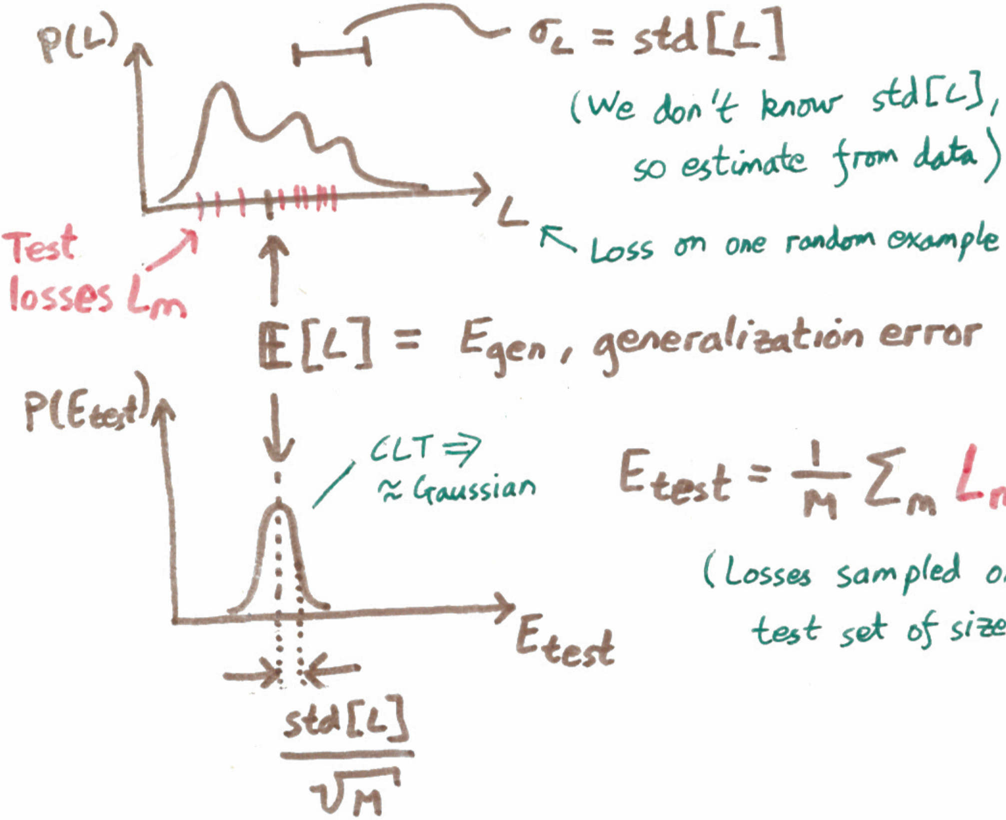
Standard error on mean test error



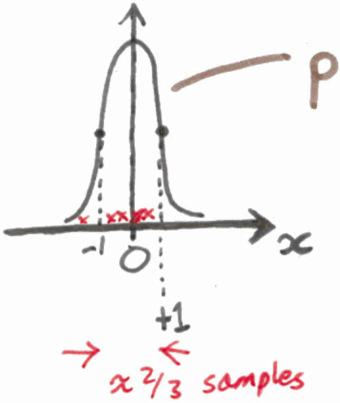
$$E_{test} = \frac{1}{M} \sum_m L_m$$

(Losses sampled on test set of size M)

⇒ For a particular fitted model,

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

UNIVARIATE GAUSSIAN REMINDER



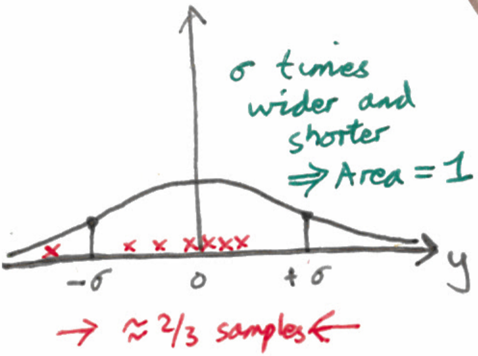
$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

$$\Downarrow$$

TRANSFORM:

$$y = \sigma x, \quad x = \frac{y}{\sigma}$$



$$p(y) = \frac{1}{\underbrace{\sigma}_{\text{scaling}}\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}$$

How variable is performance?

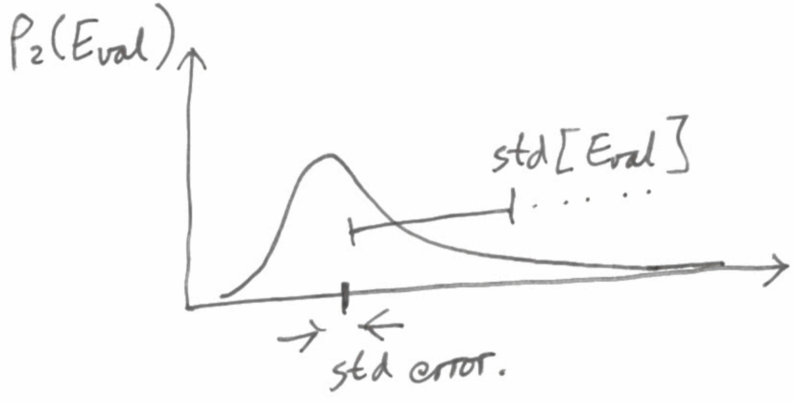
How uncertain are we about E_{gen}?

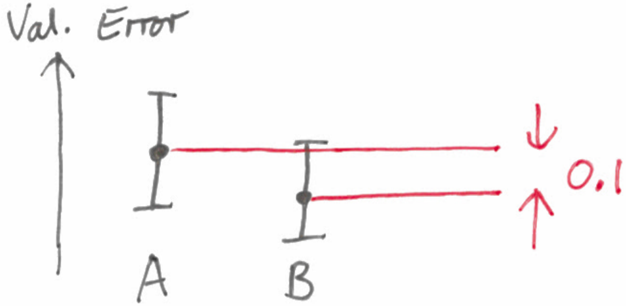
Or variability:

- Across different initialization.
- Floating point non-determinism
  - → Because of parallelism
- Use different training data

...

P_2: Distribution over E_{val} if I re-run my code





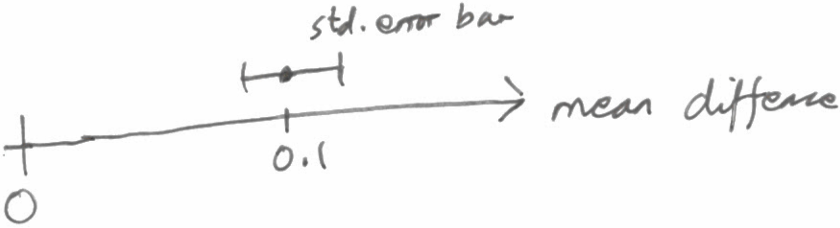
Q) Is B better than A?

Paired Comparison

Difference on example m $\delta_m = L_m^{(A)} - L_m^{(B)}$

Mean difference $= \frac{1}{M} \sum_m \delta_m$

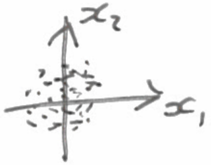
Standard error : $\frac{\text{std}[\delta_m]}{\sqrt{M}}$



Multivariate Gaussians

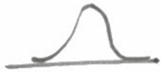
Sample $x_d \sim N(0,1)$ independently $d=1\ldots D$

(with randn)



$$p(\underline{x}) = \prod_d p(x_d)$$

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



$$= \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi}} e^{-x_d^2/2}$$

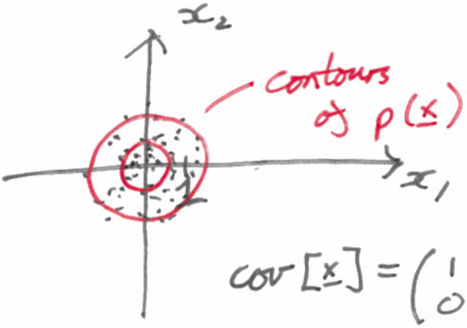
sum, not
Σ covariance

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

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

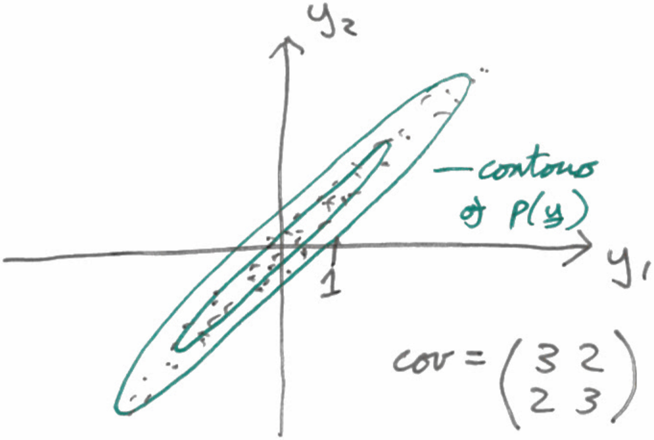
identity.

$$= N(\underline{x}; \underline{0}, \mathbb{I})$$



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

$$y^{(n)} = A\,\underline{x}^{(n)}$$



$$\mathrm{cov} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

$$= AA^T = \Sigma$$

Covariance generalization of variance

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

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

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

$$\text{cov}[\underline{x}] = \mathbb{E}[(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})^T]$$

$$\text{cov}[y] = \mathbb{E}[y y^T]$$

$$= \mathbb{E}[A\underline{x}(A\underline{x})^T]$$

$$= \mathbb{E}[A\underline{x}\,\underline{x}^T A^T]$$

$$= A\,\underbrace{\mathbb{E}[\underline{x}\,\underline{x}^T]}_{\mathbb{I} = \text{cov}[\underline{x}]}\,A^T$$

because $\underline{x}$ is zero mean

$$= AA^T = \Sigma$$

PDF of y $N(y; 0, \Sigma)$

$y = A\underline{x}$

$\underline{x} = A^{-1}y$

(If A^{-1} exists)

$$p(y) \propto e^{-\frac{1}{2}(A^{-1}y)^T(A^{-1}y)}$$

$$\propto e^{-1/2\, y^T \underbrace{A^{-T}A^{-1}} y}$$

$$\Sigma = AA^T$$

precision: $\Sigma^{-1} = A^{-T}A^{-1}$

$$\propto e^{-1/2\, y^T \Sigma^{-1} y}$$

$$p(y) = \frac{1}{|A|\,(2\pi)^{D/2}}\, e^{-1/2\, y^T \Sigma^{-1} y}$$

volume increases by $|A|$

$$|\Sigma| = |AA^T| = |A||A^T| = |A|^2$$

$$|A| = |\Sigma|^{1/2}$$

$$= \frac{1}{|\Sigma|^{1/2}\,(2\pi)^{D/2}}\, e^{-1/2\, y^T \Sigma^{-1} y}$$