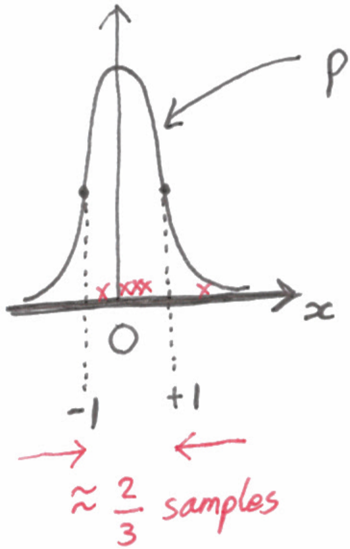
Univariate Gaussian Reminder

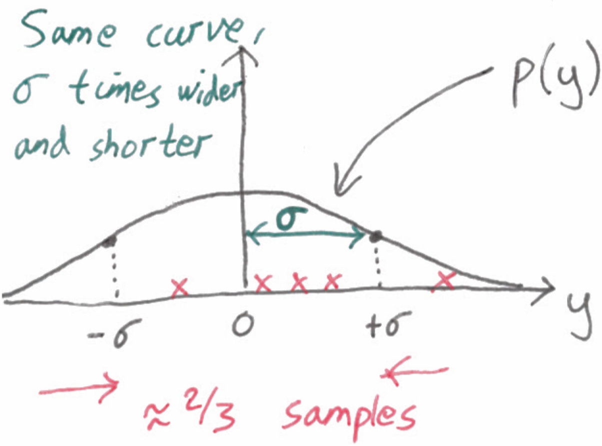


$$p(x) = \mathcal{N}(x; 0, 1)$$
$$= \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$
$$= 1$$

Transform

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



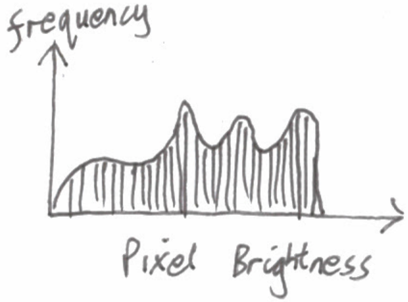
$$p(y) = \frac{1}{\underline{\underline{\sigma}}\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}}$$

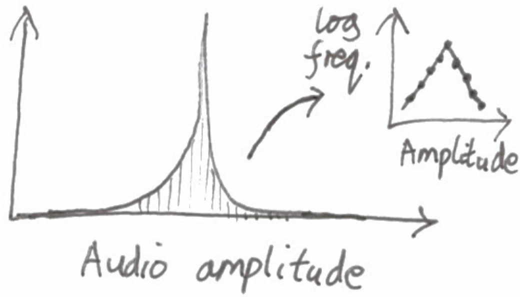
Scaling
"Jacobian of transformation"

Not every distribution is Gaussian

Can try to measure mean μ, std. dev. σ

Often $\approx 2/3$ samples <u>not</u> within $\mu \pm \sigma$





Central Limit Theorem (CLT)

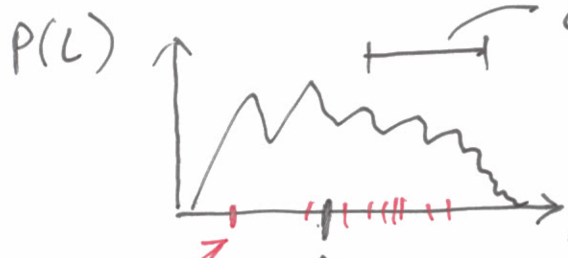
If x is a sum of
N (many)
independent outcomes
each with finite mean and variance

$x \rightarrow$ Gaussian, as $N \rightarrow \infty$

Convergence is "Convergence in distribution"

$\Rightarrow$ Don't trust Gaussian fit in tails.

Estimating generalization error



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

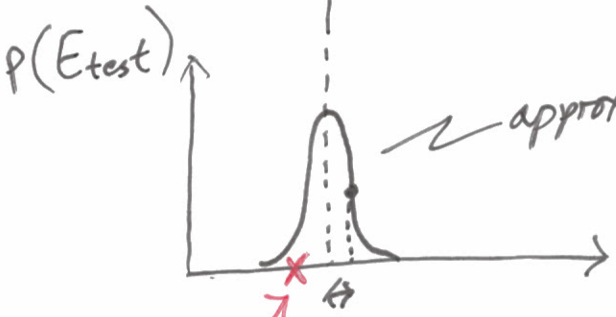
$\approx \hat{\sigma}_L$, estimate from test losses

L, Loss on one random example

M Test losses, L_m

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

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



approx Gaussian if L_m independent.

We see one sample, our test error

$\text{std}[E_{\text{test}}]$

$\Rightarrow$ Measure of how wrong we might be

$$\mathrm{var}[E_{test}] = \frac{1}{M^2} \sum_m \mathrm{var}[L_m]$$

(If test cases independent)

$$= \frac{1}{M^2} \sum_{m=1}^{M} \mathrm{var}[L]$$

$$= \cancel{\frac{1}{M^2} M} \; \frac{1}{M} \mathrm{var}[L]$$

$$\Rightarrow \mathrm{std}[E_{test}] = \frac{\mathrm{std}[L]}{\sqrt{M}} \approx \frac{\hat{\sigma}_L}{\sqrt{M}}$$

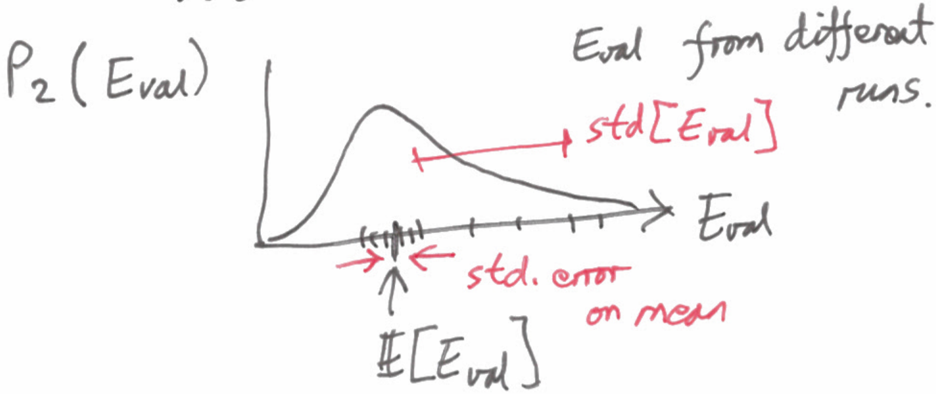
$$E_{gen} = E_{test} \pm \underbrace{\frac{\hat{\sigma}_L}{\sqrt{M}}}_{\text{Standard error on the mean.}}$$

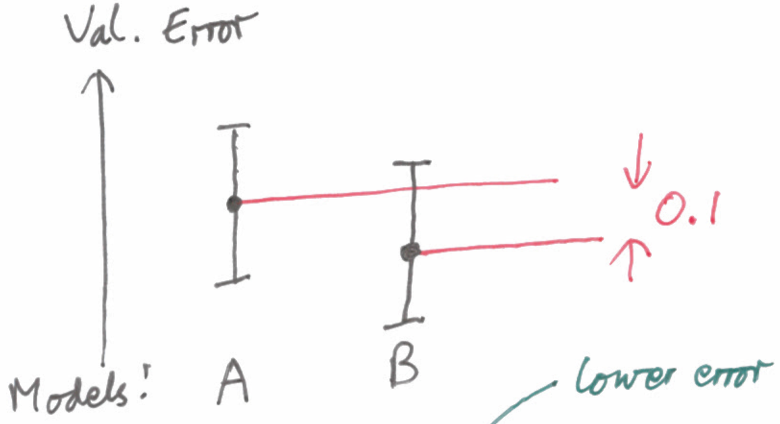
How variable is performance?

Sources of variability:

- Across different initialization or random choices.
- Floating point non-determinism
- Use different data

...





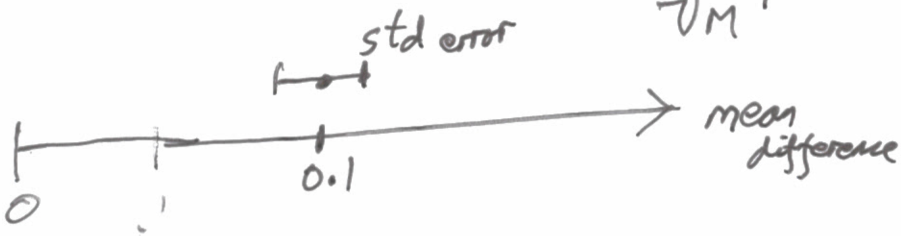
Q) Is B better than A?

<u>Paired Comparison</u>

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

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

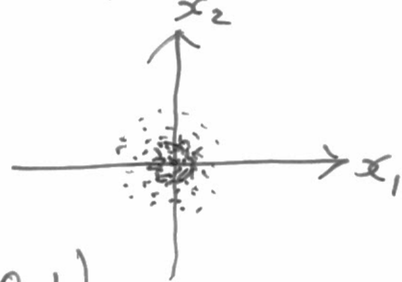
Standard error : $\dfrac{\mathrm{std}[\delta_m]}{\sqrt{M}}$



Multivariate Gaussian

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

```
xx = randn(D,1)
np.random.randn(D)
```



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

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

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

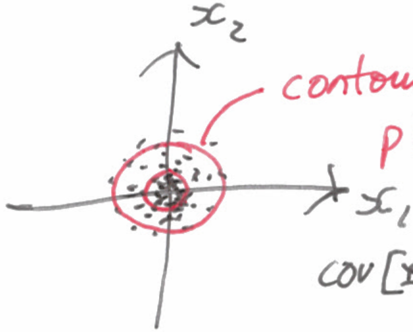
sum, not
Σ sigma
covariance

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

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

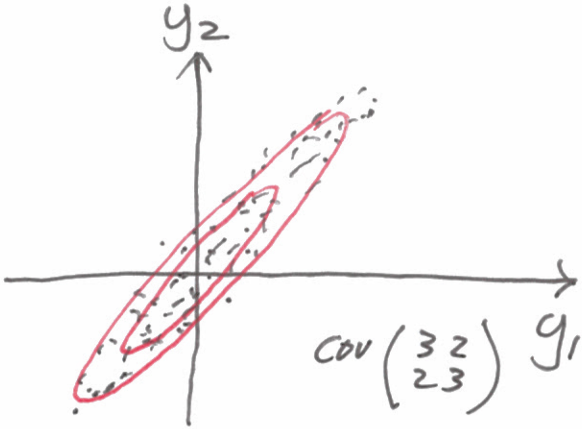
Identity

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



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

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



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

$$\mathbb{E}[\underline{y}] = \mathbb{E}[A\underline{x}] = A\underbrace{\mathbb{E}[\underline{x}]}_{\underline{0}} = \underline{0}$$

Covariance generalization of variance

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

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

$$\mathrm{cov}[\underline{x}] = \mathbb{E}[\underbrace{\underline{x}}_{D\times 1}\,\underbrace{\underline{x}^\top}_{1\times D}] - \mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]^\top$$

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

$$\mathrm{cov}[\underline{y}] = \mathbb{E}[\underline{y}\,\underline{y}^\top] \cancel{-\,0}$$

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

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

$$= A\,\underbrace{\mathbb{E}[\underline{x}\,\underline{x}^\top]}_{\mathrm{cov}[\underline{x}] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\,A^\top$$

$$= AA^\top = \Sigma, \text{ covariance of } \underline{y}$$