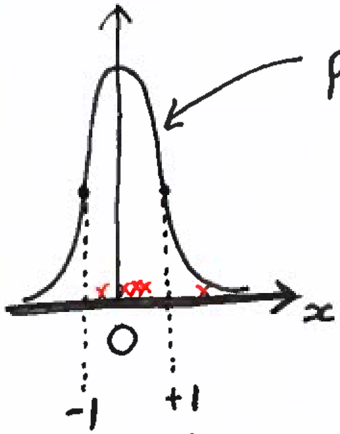
Univariate Gaussian Reminder



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

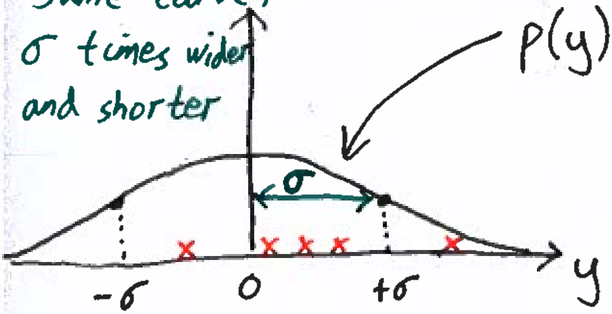
$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$
$$= 1$$

$\approx \frac{2}{3}$ samples

$\Downarrow$ **Transform**

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

Same curve, σ times wider and shorter



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

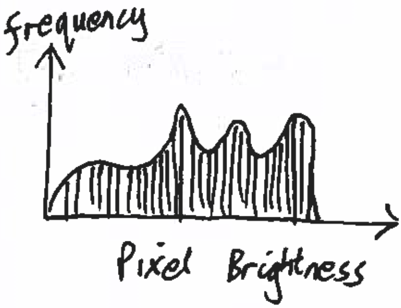
Scaling
"Jacobian of transformation"

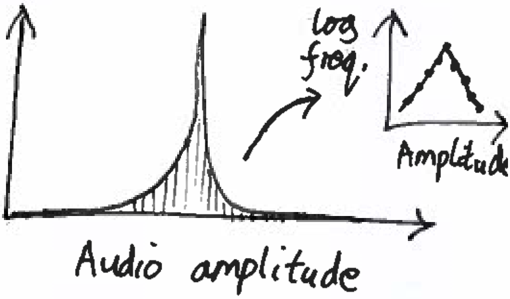
$\approx 2/3$ samples

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

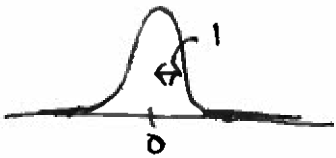
$x \to$ Gaussian, as $N \to \infty$

Convergence is "Convergence in distribution"

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

Approximating randn with 12 draws from rand

```
octave:1> xx = sum(rand(1e6, 12), 2) - 6.0;
octave:2> mean(xx)
ans =  0.0011525
octave:3> std(xx)
ans =  0.99984
octave:4> hist(xx, 1000);
```



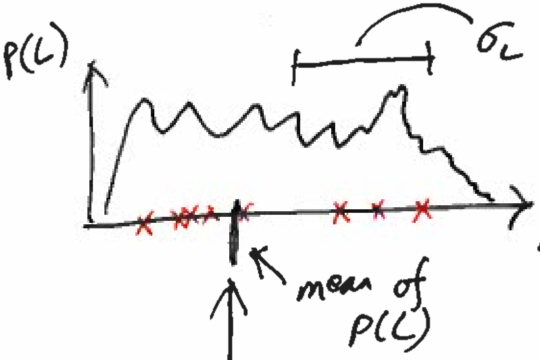

```
In [1]: xx = np.random.rand(int(1e6), 12).sum(1) - 6.0

In [2]: xx.mean()
Out[2]: 0.00015316318828693392

In [3]: xx.std()
Out[3]: 0.99959157256611364

In [4]: plt.hist(xx, 1000); plt.show()
```

Estimating generalisation error



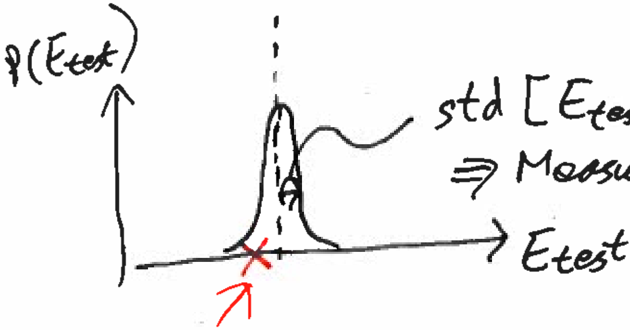
$\sigma_L = \mathrm{std}[L]$

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

Loss on one example

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

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



std $[E_{test}]$

$\Rightarrow$ Measure of how wrong we'll be

We see <u>one</u> sample, our test error

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

(If test cases independent)

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

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

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

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

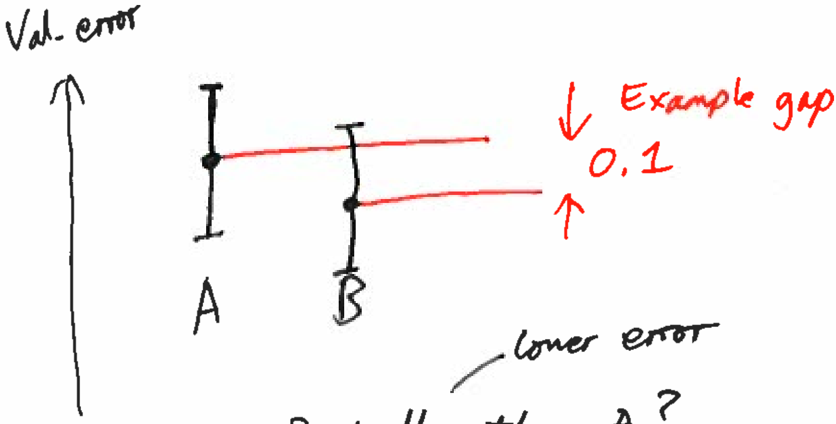
How variable is performance?

Sources of variability:

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

...

$P_z(E_{val})$, dist due to randomness



lower error

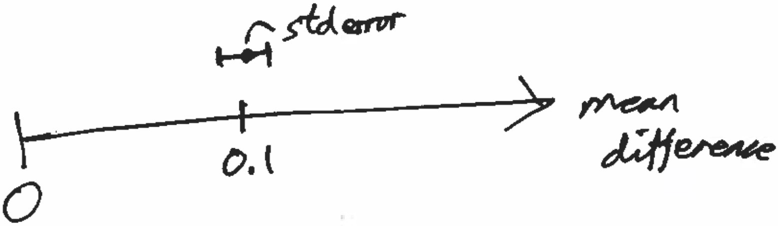
Q Is B better than A?

Paired Comparison

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

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

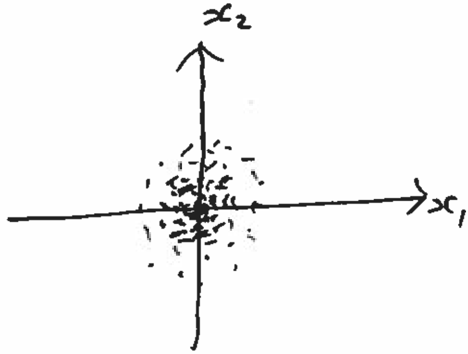
Standard error: $\quad \dfrac{\mathrm{std}[\delta_m]}{\sqrt{M}}$



Multivariate Gaussian

Sample $x_d \sim \mathcal{N}(0,1)$, independently $d=1 \dots D$

```
X = randn(N, D);
   ^
np.random.
```



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

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

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

sum Σ_1^d not a Sigma Σ

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

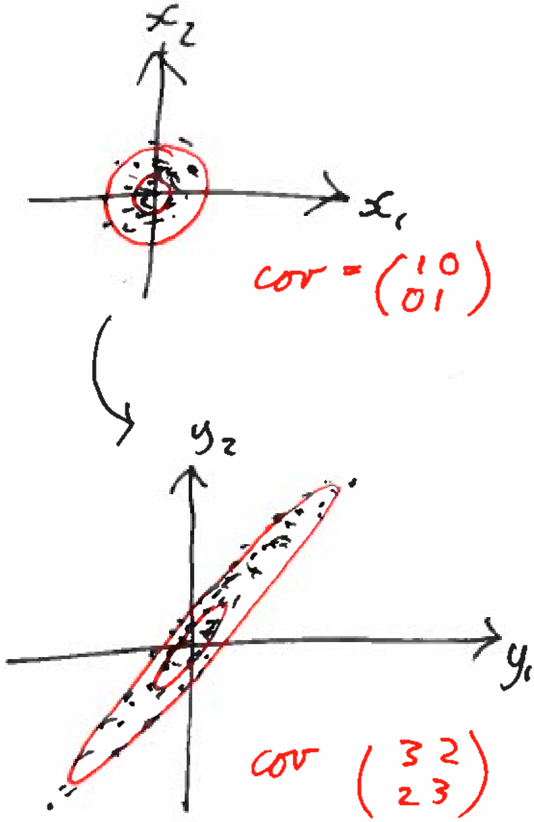
$$= \frac{1}{(2\pi)^{D/2}} e^{-1/2\, \underline{x}^\top \underline{x}}$$

Identity

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

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

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



cov

Covariance generalization of variance

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

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

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

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

$$\text{cov}[\underline{y}] = \mathbb{E}[\underline{y}\underline{y}^T] \cancel{- 0}$$

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

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

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