Univariate Gaussian Reminder

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

\[ \text{var}[x] = E[x^2] - E[x]^2 = 1 \]

\[ \text{Transform} \quad y = \sigma x, \quad x = \frac{y}{\sigma} \]

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

Scaling "Jacobian of transformation"

Same curve, \( \sigma \) times wider and shorter

\( \sim \frac{2}{3} \) samples

\( \sim \frac{2}{3} \) samples
Not every distribution is Gaussian

Can try to measure mean $\mu$, std. dev. $\sigma$

Often $\approx \frac{2}{3}$ samples not 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

\[ p(L) \]

\[ \sigma_L = \text{std}[L] \]
\[ \approx \hat{\sigma}_L, \text{ estimate from test losses} \]
\[ L, \text{ Loss on one random example} \]

\[ E[L] = E_{\text{gen}}, \text{ generalization error} \]
\[ \approx E_{\text{test}} = \frac{1}{M} \sum_{m} L_m \]

\[ p(E_{\text{test}}) \]

\[ \text{approx Gaussian if } L_m \text{ independent.} \]

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

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

\[ \Rightarrow \text{Measure of how wrong we might be} \]

We see one sample, our test error
\[ \text{Var}[E_{test}] = \frac{1}{M^2} \sum_{m=1}^{M} \text{Var}[L_m] \]

\[ \text{If test cases independent} \]

\[ = \frac{1}{M^2} \sum_{m=1}^{M} \text{Var}[L] \]

\[ = \frac{1}{M^2} M \frac{1}{M} \text{Var}[L] \]

\[ \Rightarrow \text{std}[E_{test}] = \frac{\text{std}[L]}{\sqrt{M}} \approx \frac{\hat{s}_E}{\sqrt{M}} \]

\[ E_{gen} = E_{test} \pm \frac{\hat{s}_E}{\sqrt{M}} \]

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

\[ P_2(\text{Eval}) \]

Eval from different runs.

\[ \text{std}[\text{Eval}] \]

\[ \text{std. error on mean \[ \text{E}[\text{Eval}] \]} \]
Val. Error

Models: A B

lower error

Q) Is B better than A?

Paired Comparison

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

Mean difference = \( \frac{1}{M} \sum_m S_m \)

Standard error: \( \frac{\text{std}[S_m]}{\sqrt{M}} \)

mean difference
**Multivariate Gaussian**

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

$$x \mathbf{k} = \text{randn}(D, 1)$$

$$n \mathbf{p} \cdot \text{random} \cdot \text{randn}(D)$$

$$p(x) = \prod_d p(x_d)$$

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

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

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

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

$$= N(x; 0, \mathbb{I})$$
\[ y^{(n)} = A x^{(n)} \]

\[ E[y] = E[Ax] = A E[x] = 0 \]

\[ \text{Cov} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \]
Covariance generalization of variance

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

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

$\text{cov}[x] = E[x x^T] - E[x] E[x]^T$

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

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

$= E[(A x)(A x)^T]$

$= E[A x x^T A^T]$

$= A E[x x^T] A^T$

$\text{cov}[x] = \Sigma$, covariance of $y$

$\text{cov}[x] = \Sigma$, covariance of $y$