tinyurl.com/edmlpr

First tutorial sheet is up
First tutorial next week
Answers released end of next week

Hypothesis Forum
- Share links, code
- Get code review
- Ask questions
- Post answers
  - help others
  - check if right
Generalization

\[ E_{gen} = \mathbb{E}_{p(x,y)} \left[ L(y, f(x)) \right] \]

\approx \frac{1}{M} \sum_{m=1}^{M} L(y^{(m)}, f(x^{(m)})) = E_{test} \]

\[ x^{(m)}, y^{(m)} \sim p(x, y); \quad M \text{ held-out data examples} \]

Need model choice and \{x^{(m)}, y^{(m)}\} independent.

\Rightarrow \text{Model not chosen using } E_{test}

Assume there is a fixed distribution \( p(x, y) \)
How do we avoid fitting test set?
Reduce need to look at test set

**k-fold cross validation**

- Data in K Chunks
- validation
- train

Pick Model (or λ) with lowest average validation set loss
How do we deal with $p(x, y)$ changing?

$A \uparrow$ Depends

$p(x, y) = p(x)p(y|x)$ (product rule)

Input dist. Noisy mapping to outputs

When $p(x)$ changes.

If $p(y|x)$ changes? generalization

No $y$'s at test time
How would we know?
Need some information about the change.

Amos Storkey has review
Gaussian (Univariate)

Draw $10^6$ values $x \sim N(0,1)$

$\sim 68\% \text{ area within } \pm 1$

$\sim 95\% \text{ within } \pm 2$

$$z = \sigma x + \mu$$

$$x = \frac{z - \mu}{\sigma}$$
Area should be the same.

Height decreased by $\sigma \sim 68\%$ area within $\pm \sigma$

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

\[ p(z) = N\left( \frac{z}{\sigma}; \mu, \sigma^2 \right) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left( \frac{z - \mu}{\sigma} \right)^2 / 2} \]

\[ \text{variance} = \sqrt{2\pi \sigma^2} \]
Not every distribution is Gaussian

Probability mass within ±6 ≠ 68% in general

e.g. Load audio file, histogram of amplitudes

\[ \text{Log scale} \]

Central Limit Theorem (CLT)

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

\[ x \rightarrow \text{Gaussian as } N \rightarrow \infty \]

\[ x = \frac{1}{N} \sum_{n=1}^{N} q_n \]

Convergence is convergence by distribution

Don't trust Gaussian fit in the tails.
Error bars

\[ E_{\text{test}} = \frac{1}{m} \sum_m L_m \]

loss on mth example

\[ E \left[ E_{\text{test}} \right] = E_{\text{gen}} \]

\[ \text{var} \left[ E_{\text{test}} \right] = \frac{1}{M^2} \sum_m \text{var} [L_m] \]

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

\[ = \frac{\text{var} [L]}{M} \]

\[ \text{std} \left[ E_{\text{test}} \right] = \frac{\sqrt{\text{var} [L]}}{\sqrt{M}} \]

\[ = \frac{\text{std} [L]}{\sqrt{M}} \]

\[ E_{\text{gen}} = E_{\text{test}} \pm \frac{\text{std} [L]}{\sqrt{M}} \]