tinyurl.com/edmlpr

Sign up for a tutorial group

Do the tutorial sheet!

~200 of you are on Hypothesis

... the rest: please sign up

Duplicate Lectures

Today: AT LT5 2:10pm

Thur: DHT LTC 4:10pm
(not Fri)

(Next week TBA.)

Generalization

Loss function

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

$$\approx \frac{1}{M} \sum_{m=1}^{M} L(y^{(m)}, f(\underline{x}^{(m)}))$$

$$= E_{test}$$

Assume: M held-out test examples $\underline{x}^{(m)}, y^{(m)} \sim p(\underline{x}, y)$

Model $f()$ and $\{\underline{x}^{(m)}, y^{(m)}\}$ independent

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

Validation / Development set(s) to make choices

Eg, fit $\underline{w}$ on training set for $\lambda = 0.1, 1, 10,$

Pick from these 3 models using validation loss

$\rightarrow$ "Fit λ to validation set"

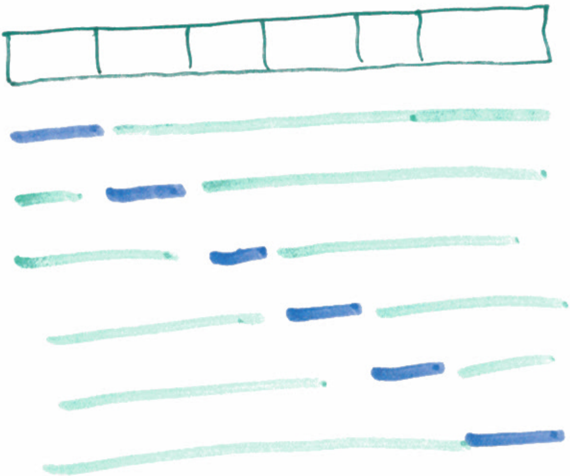
How do we avoid fitting test set?

- Reduce amount look at validation or test sets.

- More than one validation set?

K-fold cross validation

Train data



Data in K chunks.

K is often ≈10

validation

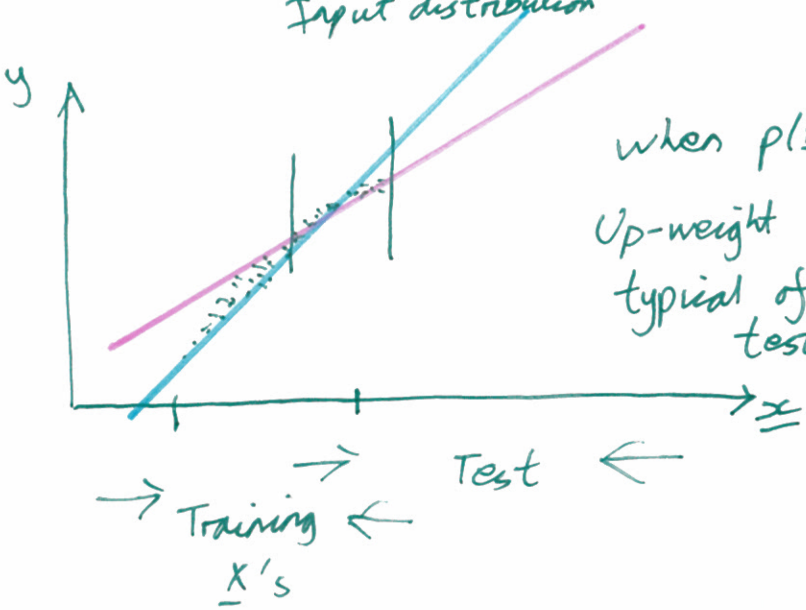
training

Pick λ or model based on average of validation scores

How do we deal with $p(\underline{x}, y)$ changing?

Answer: it depends.

$$p(\underline{x}, y) = \underbrace{p(\underline{x})}_{\text{Input distribution}} \overbrace{p(y \mid \underline{x})}^{\text{Noisy mapping between inputs \& outputs}} \quad \text{(product rule)}$$



when $p(\underline{x})$ changes.

Up-weight $\underline{x}$'s more typical of test time.

what $p(y \mid \underline{x})$ changes?

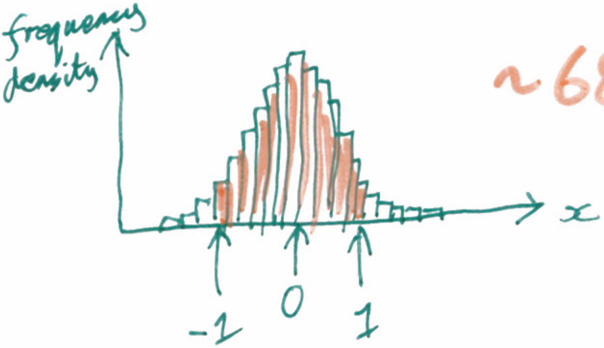
Might not get new labels in future.

Need some information about change.

Amos Storkey has review

Gaussian (Univariate)

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



~68% of area within ± 1

~95% is within ± 2

$z = \sigma x + \mu$

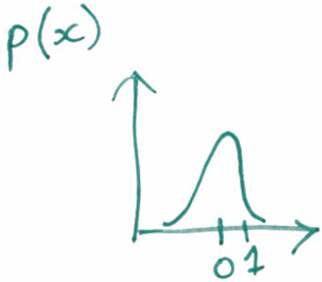
$x = \frac{z-\mu}{\sigma}$

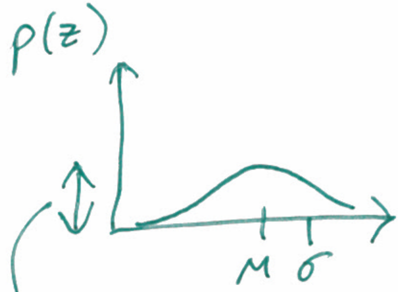
variance of z points is σ^2

Mean is μ.

$z \sim N(\mu, \sigma^2)$

variance





height is $\frac{1}{\sigma} \cdot$ [height of $p(x)$]

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

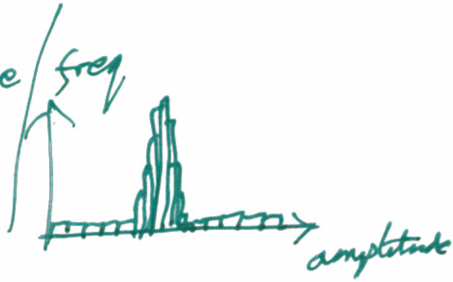
$$p(z) = N(z; \mu, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma}}_{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$$

Not every distribution is Gaussian

Prob. mass within $\pm\sigma \neq 68\%$

eg Load an audio file

histogram amplitude



Central Limit Theorem (CLT)

If x is a sum of

N (many)

independent outcomes,

each with finite mean, finite variance.

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

Convergence is "convergence in distribution"

Don't trust Gaussian fit in tails.

Error bars

$$E_{test} = \frac{1}{M}\sum_{m=1}^{M} L_m$$

loss on m^{th} example.

$$\mathbb{E}[E_{test}] = E_{gen}$$

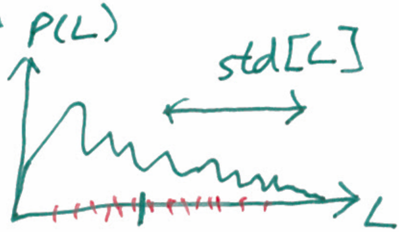
$$\text{var}[E_{test}] = \frac{1}{M^2}\sum_{m=1}^{M} \text{var}[L_m]$$

(test cases independent)

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

$$= \frac{1}{M}\text{var}[L]$$

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



Standard error on the mean

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