tinyurl.com/edmlpr

- ITO is organizing tutorial groups
- Do the tutorial sheet

~70 of you not on Hypothesis.
Please sign up.

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

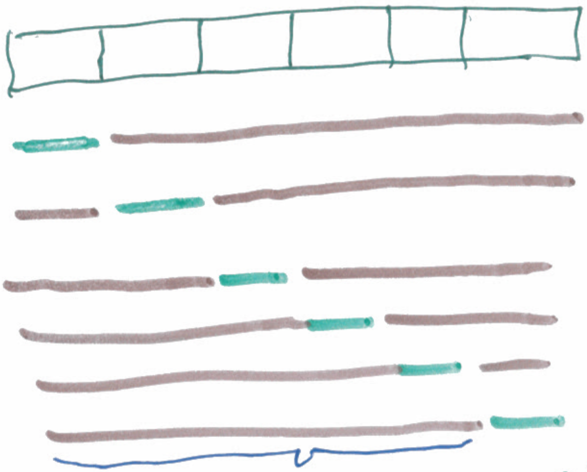
$\rightarrow$ "Fit λ to validation set"

How do we avoid fitting test set?

- Reduce amount look at validation set or test set.

- More than one validation set?

k-fold cross validation



Train data split into K chunks

Validation

training

Here k=6

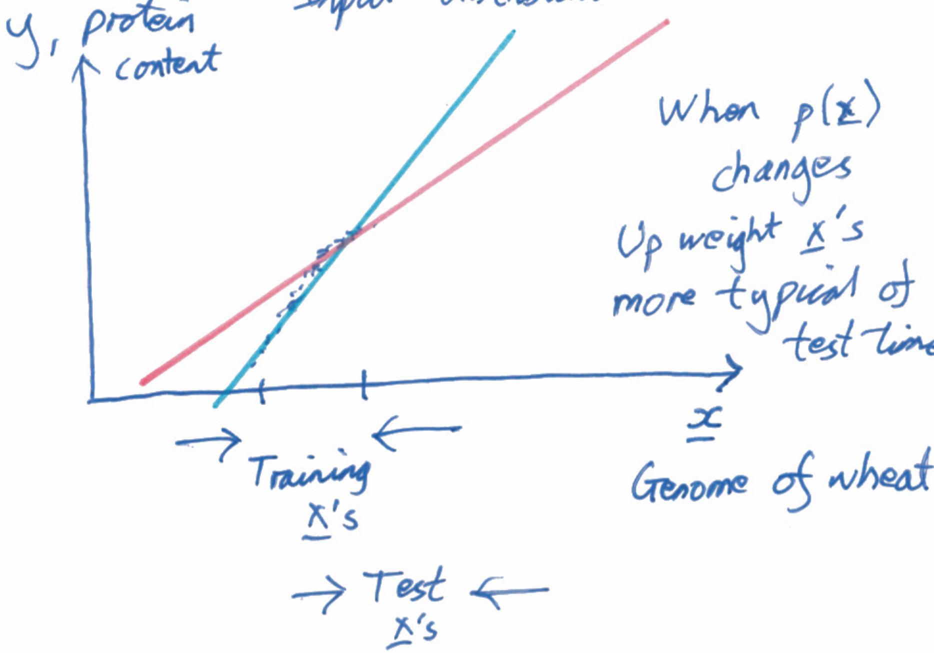
Pick λ or model based on average validation score.

$\boxed{6}$ $\boxed{0}$ ... $\boxed{1}$ $\boxed{1}$ ... $\boxed{9}$

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

Answer: it depends

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



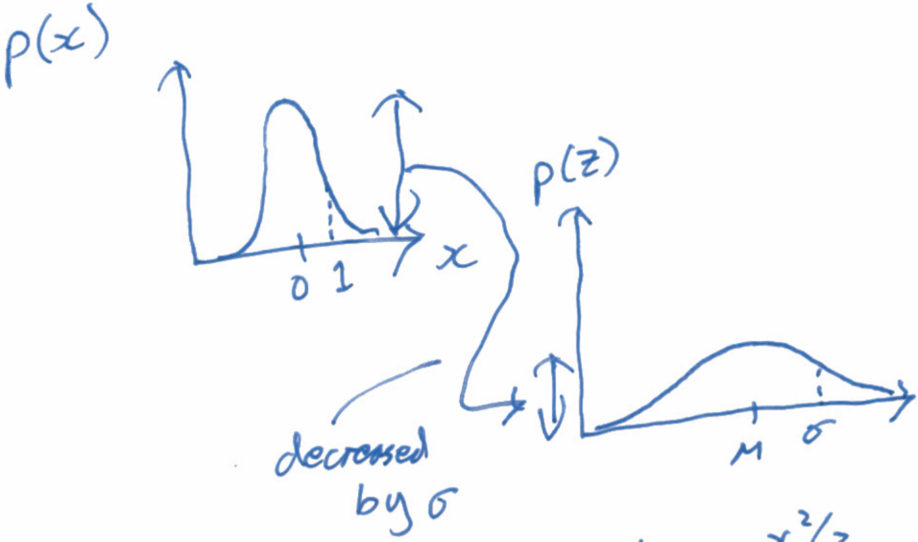
When $p(\underline{x})$ changes
Up weight $\underline{x}$'s more typical of test time

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

Might not get new labels in future?

Need some information about change...

Amos Storkey has review

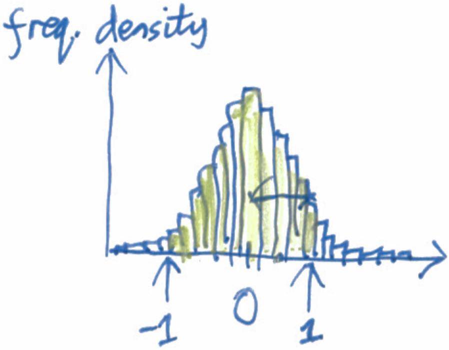


$$p(x) = \mathcal{N}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

Gaussian (Univariate)

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



~68% of area within ±1

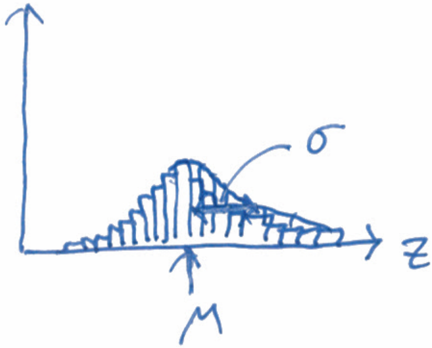
~95% within ±2

$$z_n = \sigma x_n + \mu$$

$$x_n = \frac{z_n - \mu}{\sigma}$$

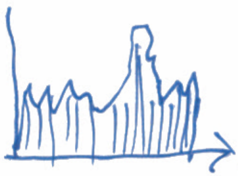
Variance of z points is σ^2

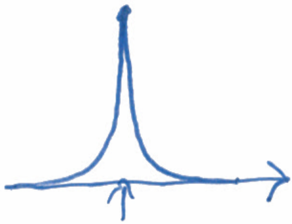
Mean of points is μ



$$z_n \sim \mathcal{N}(\mu, \underset{\text{variance}}{\sigma^2})$$

Not every distribution is Gaussian





Central Limit Theorem (CLT)

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

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