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

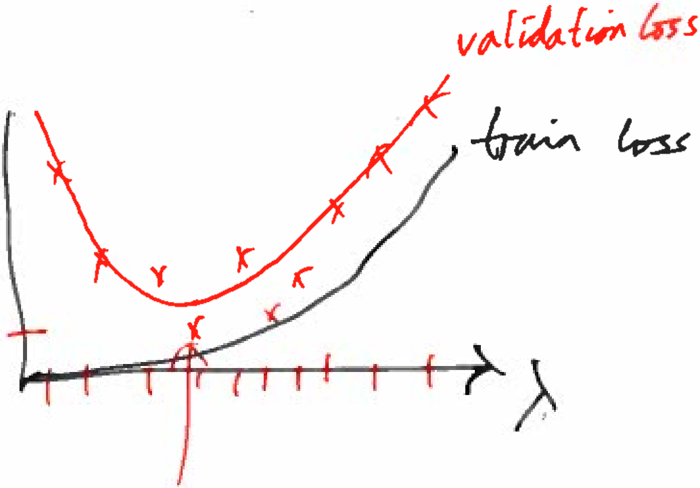
$\Rightarrow$ Model not chosen using E_{test}

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"

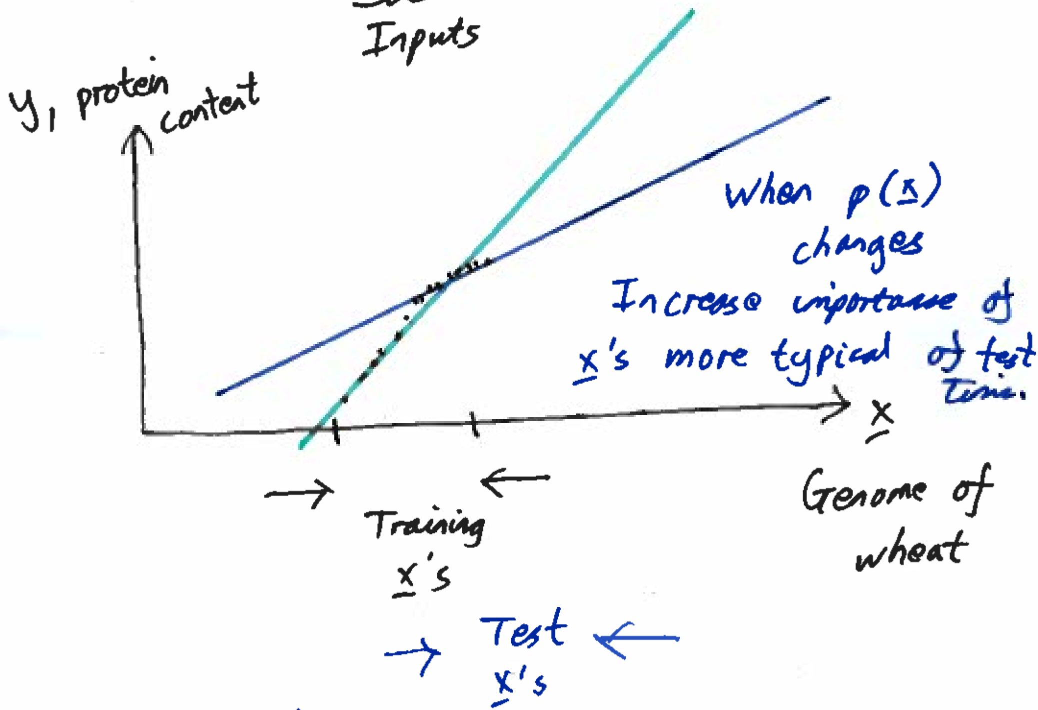


2019 L5 2

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

Answer: it depends

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



When $p(\underline{x})$ changes

Increase importance of $\underline{x}$'s more typical of test time.

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

Need some information about change...

More data in future, or more "inputs" to model, eg. time

Amos Storkey has review

How do we avoid fitting test set?

- Reduce amount we look at test set
  also at validation set.
  or have huge validation set
- More than one validation set?

k-fold cross validation

Train data split into K pieces.

validation

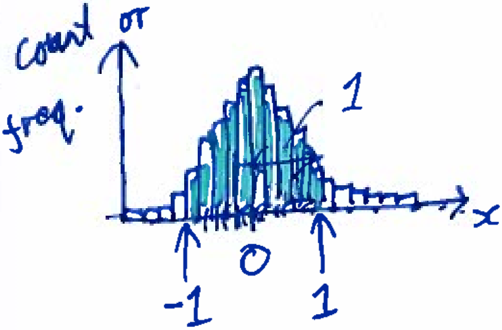
training

Average validation scores across folds

Gaussian (Univariate)

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

mean



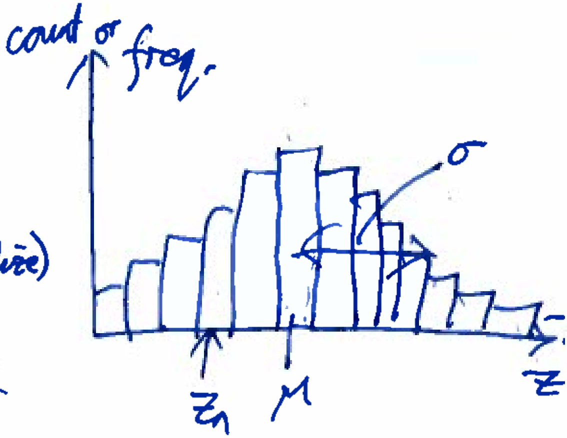
~ 68% of area within ±1

~95% within ± 2

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

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

(standardize)

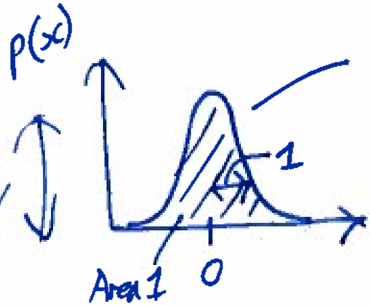


Variance of z points is σ^2

Mean of points is μ

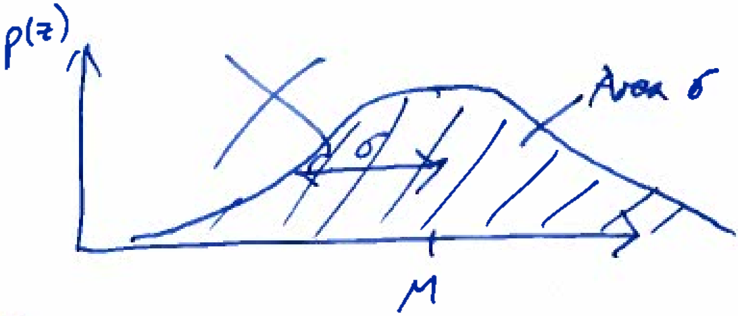
$$z_n \sim \mathcal{N}(\mu, \sigma^2)$$

variance

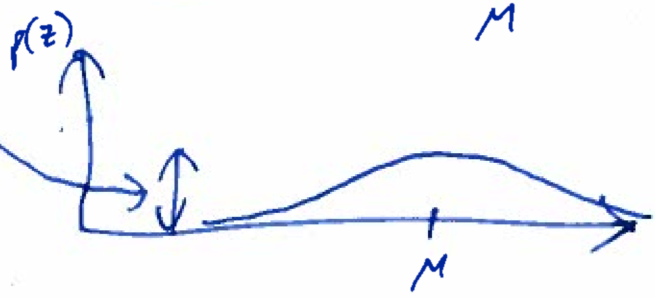


$$p(x) = \mathcal{N}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

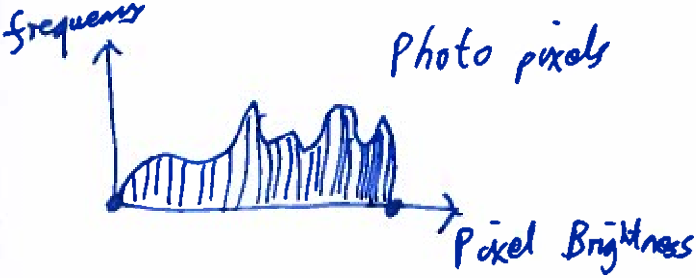


decrease by σ



"Jacobian of the transform"

Not every distribution is Gaussian



Central Limit Theorem (CLT)

If x is a sum of

N (many)

independent outcomes

each finite mean and variance

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