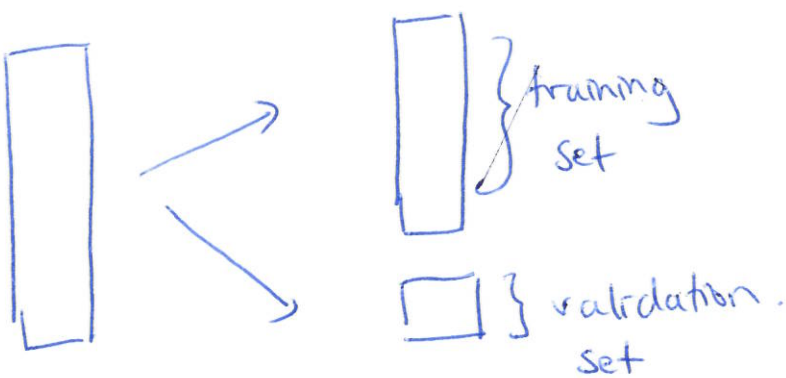
10-601

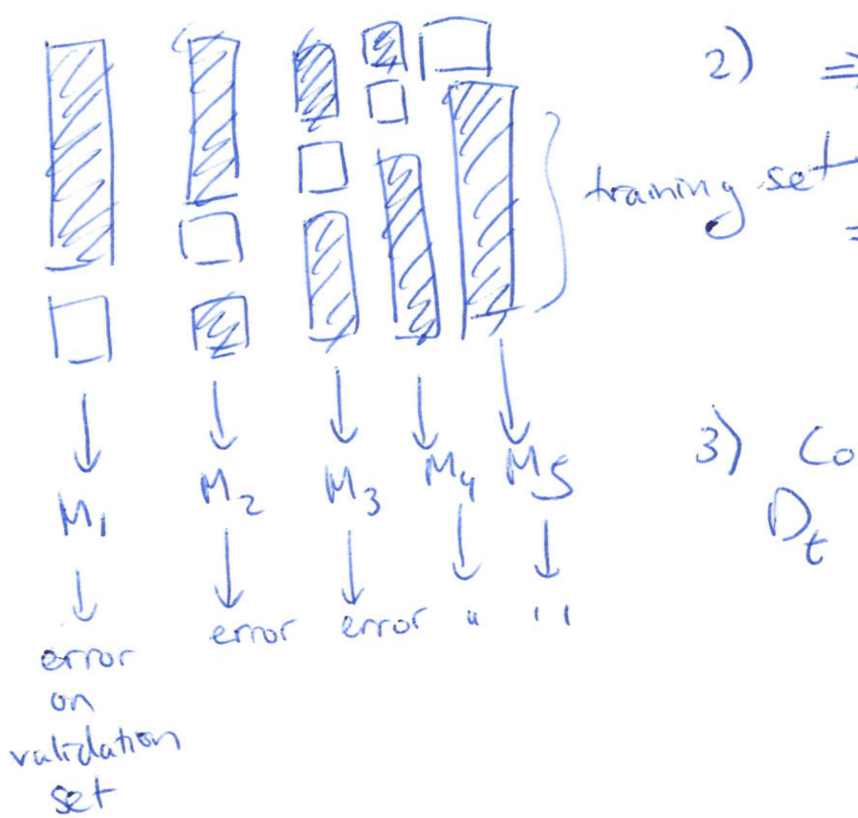
Jing Xiang

D - training data $\qquad$ D_t - test data.

K = 5



1)



2) ⇒ pick model with lowest error

⇒ take all of training data D and refit parameters

3) Compute error on test data D_t given model from 2).

unknown

$$E_Y \left[D_{KL} \left(f \parallel g(x \mid \theta_{ML}(y)) \right) \right]$$

Kullback
Liebler Divergence

- distance btw f and g.
- estimate of the information lost when a given model is used.

$$A = \underbrace{\log g(x \mid \hat{O}(y))}_{\text{likelihood}} - k$$

likelihood want to be large, meaning that we have a good fit

penalty for including extra predictors.

frequentist

- probabilities based on frequencies observed in the real world
- parameters are fixed unknown constants

$P(\nu \mid D)$ ✗

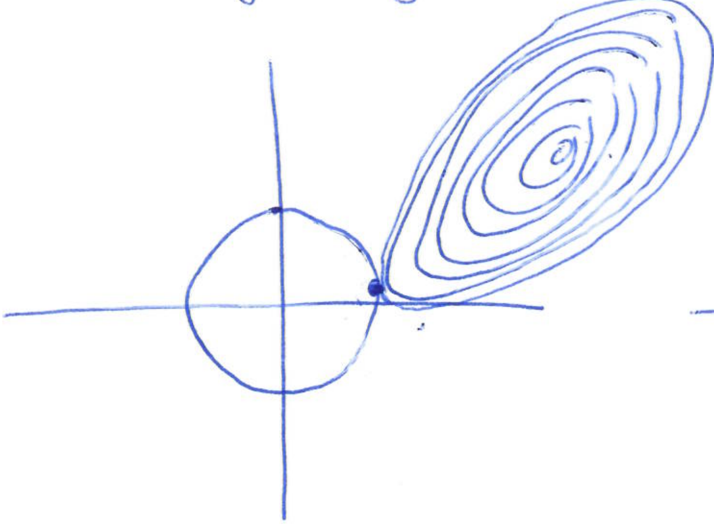
Bayesian

→ probabilities are degrees of belief.

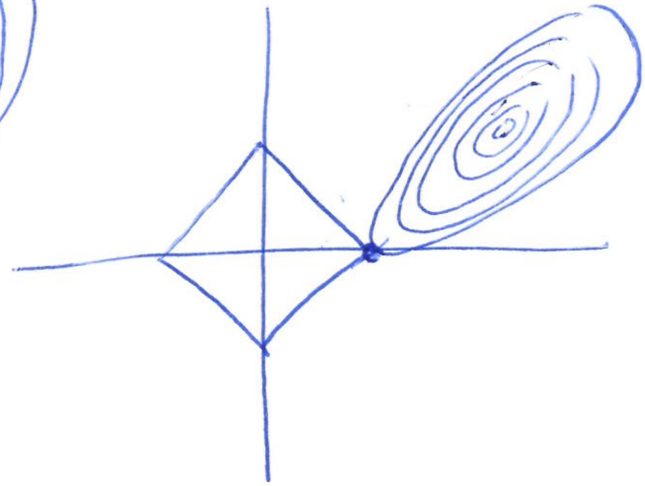
→ parameters are random variables.

$P(\nu \mid D)$ ✓

Ridge Regression



Lasso popular!



In practice to avoid biased weights:

Lasso $\longrightarrow$ β sparse

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ nz \\ 0 \\ 0 \\ nz \\ nz \end{bmatrix}$$

$\longrightarrow$ in practice, take these features and

$\longrightarrow$ then refit wil linear regression

Ridge regression

$$y = X \beta$$

trait genotype data. association strength

$$P(\Delta \mid D) = \sum_{i=1}^{I} P(\Delta \mid M_i, D)\, P(M_i \mid D)$$

quantity of interest

posterior dist of Δ given the model $\Downarrow$ different models

posterior ~~dist~~. model probability. $\Downarrow$ weighted average.

$$P(M_i \mid D) = \frac{P(D \mid M_i)\, P(M_i)}{\underbrace{P(D)}_{\text{constant}}}$$

uniform

$$= P(D \mid M_i)$$

$$= \int_{\theta} P(D \mid \theta_i, M_i)\, P(\theta_i \mid M_i)\, d\theta.$$

$$\approx \underbrace{\log P(D \mid \hat{\theta}_i) - \frac{k}{2} \log N}_{} = \text{BIC}$$

$\Downarrow$ ML.