COMP90051 **Statistical Machine Learning**

Semester 2, 2015

Lecturer: Ben Rubinstein

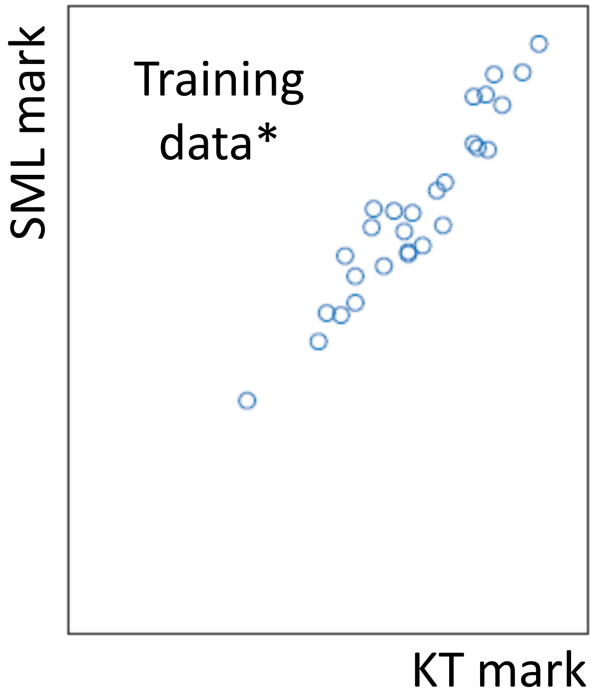
4: Extra – From Workshop #2 Feedback

THE UNIVERSITY OF MELBOURNE

Common Confusion

- What the “y axis” represents in linear model for logistic regression (log odds of probability label is True)
- Not realising that logistic regression is fit using MLE
- Linear regression has exact formula that data can be plugged into; logistic regression needs numerical methods
- Slides04.5
- How does λ control model complexity?
- Pros / cons of Lasso vs Ridge regression

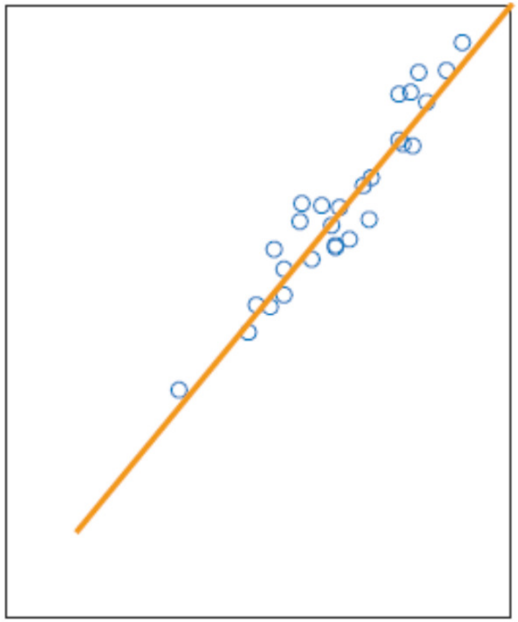
Data is noisy



- Example:
  - given mark for Knowledge Technologies (KT)
  - predict mark for Statistical Machine Learning (SML)

* synthetic data :)

Types of models



$$\hat{y} = f(x)$$

KT mark was 95, SML mark is predicted to be 95

MLE for linear/logistic regression

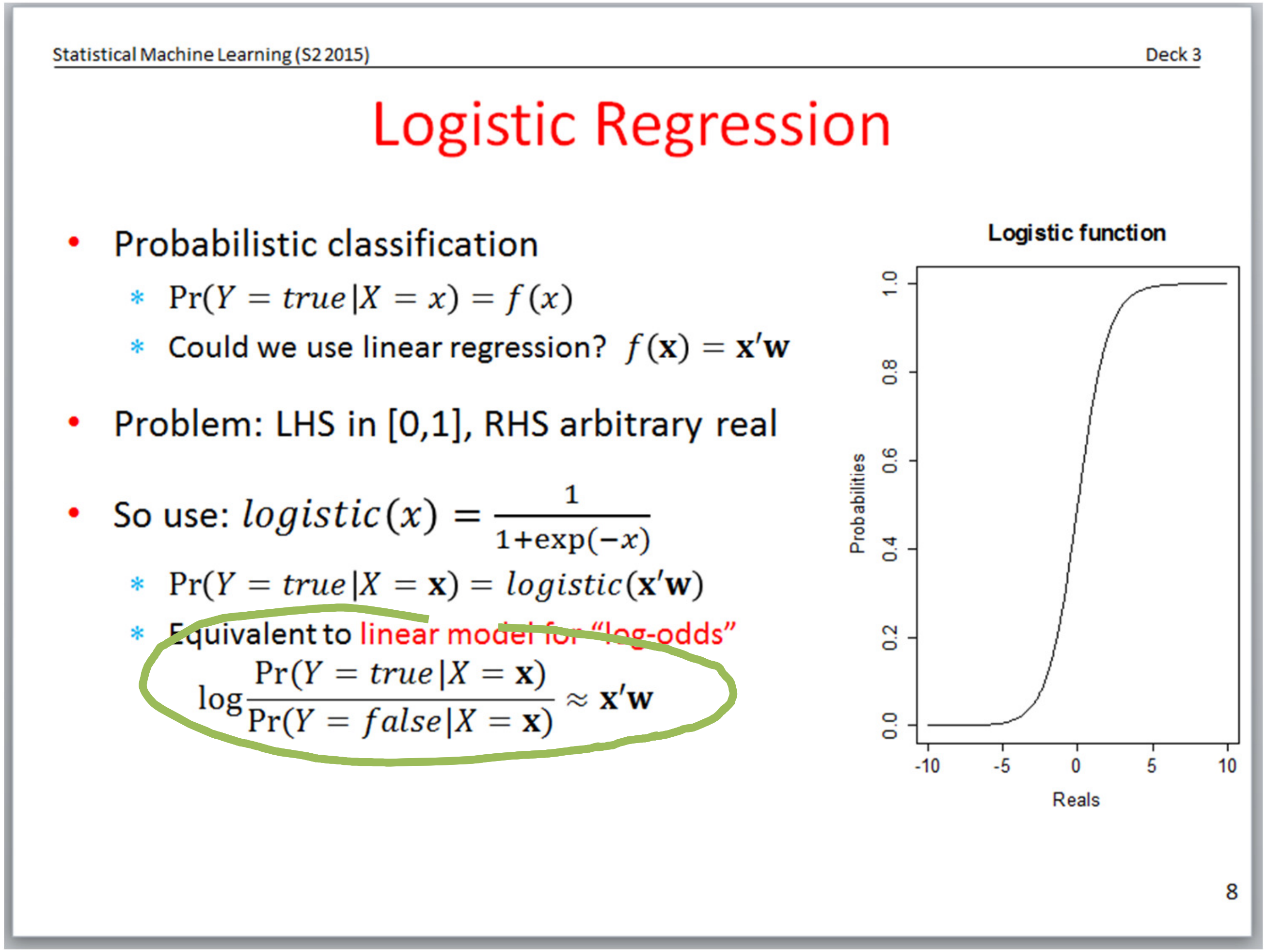
- Both have probabilistic models relating X,Y with param **w**
- Use MLE to find param **w** that says training data likely
- Linear regression
  - Model Pr(Y|X=**x**) is Normal with mean **w'x**
  - MLE gives us maximisation that is same as least squares
  - Solution has formula → just plug in data!
- Logistic regression
  - Model Pr(Y=True|X=**x**)=logistic(**w'x**)
  - MLE maximisation is an ugly one*!!*
  - Solution has no formula, use numerical approximation

How is “logistic regression” regression? What’s y-axis?

Answer:
Log odds of probability of label being True

Example: x′w=1.2

$$log_e \frac{\Pr(T)}{\Pr(F)} = 1.2$$
$$\frac{\Pr(T)}{\Pr(F)} = e^{1.2} = 3.3$$
$$\Pr(T) = 3.3(1 - \Pr(T))$$
$$\Pr(T) = \frac{3.3}{4.3} = 0.77$$



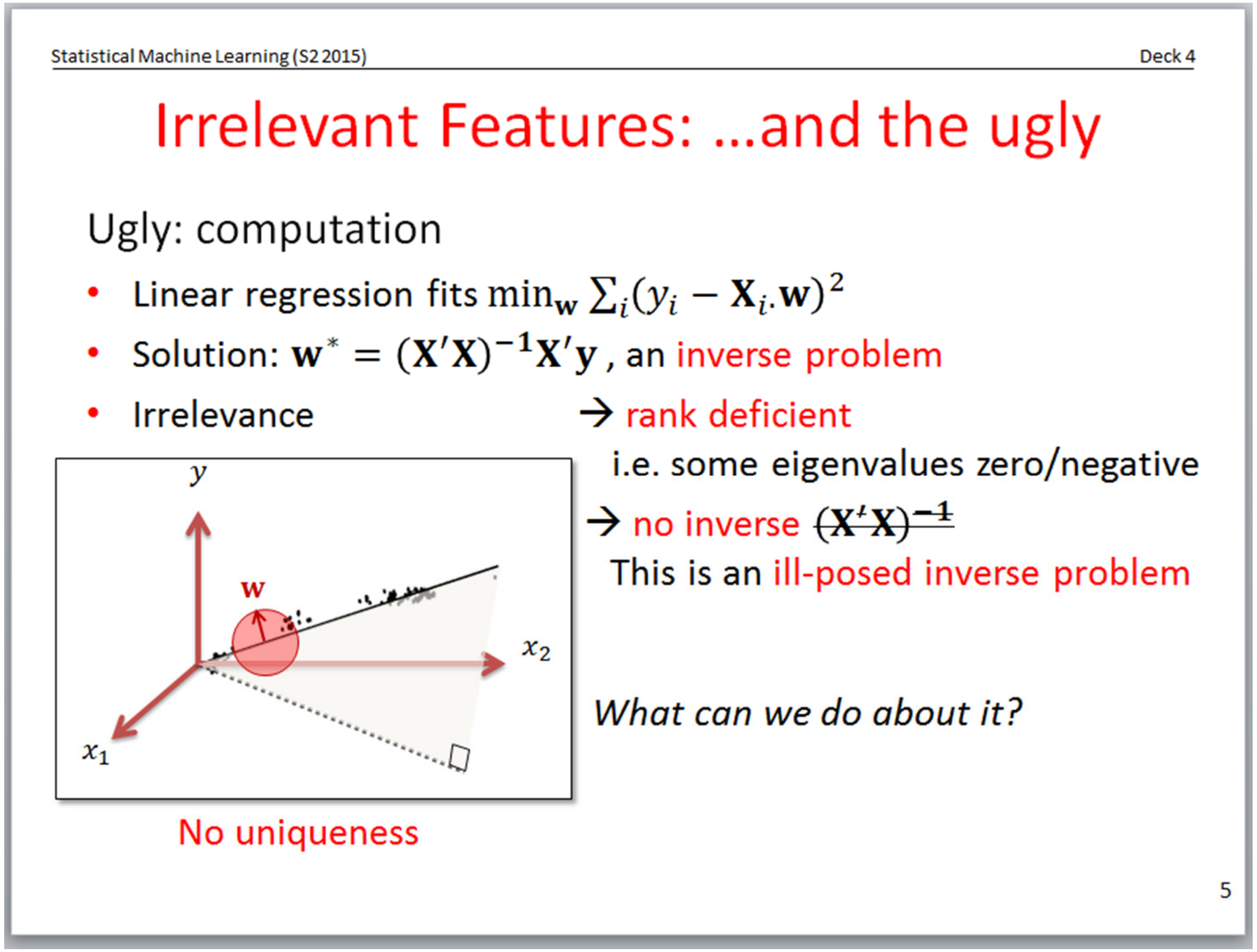
Slides04… Part I

Linear regression usually:

- Data would be spread over a plane
- Unique **w**

Irrelevant features:

- Data spread over a line
- Many planes intersect line
- Many **w**



Statistical Machine Learning (S2 2015) Deck 4

Irrelevant Features: …and the ugly

Ugly: computation

- Linear regression fits $\min_{\mathbf{w}} \sum_i (y_i - \mathbf{X}_i.\mathbf{w})^2$
- Solution: $\mathbf{w}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, an inverse problem
- Irrelevance → rank deficient

i.e. some eigenvalues zero/negative

→ no inverse $\cancel{(\mathbf{X}'\mathbf{X})^{-1}}$

This is an ill-posed inverse problem

No uniqueness

What can we do about it?

5

Slides04… Part I

Plots are top down

Pink curves are contour lines of the objective function (like a topographical map!)

Blue regions restrict where we can pick **w** from–regularisation!