tinyurl.com/edmlpr

(Try refreshing the page)

Tutorials:

- 1st sheet up
- Meetings next week
  (Groups MyEd + Learn, may have to change some...)
- Answers released end of next week

Hypothesis Forum

code

- Share links, code snippets
- Get code review
- Ask Questions
- Post Answers

help others
get feedback

Lecture theatres after today

Tuesdays 9am:

Appleton Tower LT5

Wednesdays 9am:

David Hume Tower LTs, LTA

Lecture theatres behind DHT itself

Thursdays 9am:

Appleton Tower LT5

Linear Regression Reminders

Model $f(\underline{x}) = \underline{w}^T \underline{\phi}(\underline{x})$

Can minimize

$$\underbrace{\sum_n \left(y^{(n)} - \underline{w}^T \underline{\phi}(\underline{x}^{(n)})\right)^2}_{= (y - \Phi \underline{w})^T (y - \Phi \underline{w})} \quad \text{wrt } \underline{w}$$

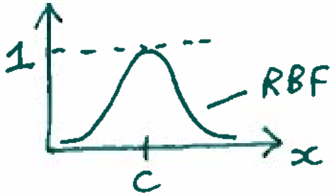
$$\underline{\phi}(\underline{x}) = [\phi_1(\underline{x}) \;\; \phi_2(\underline{x}) \; \ldots \; \phi_K(\underline{x})]^T$$

$\phi_k(\underline{x})$ any scalar function

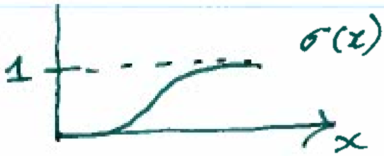
- Monomial, eg x_2, $x_3 x_4^3$, ...
- RBF

- Sigmoid

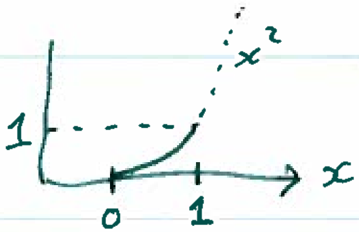
Why are large weights bad?

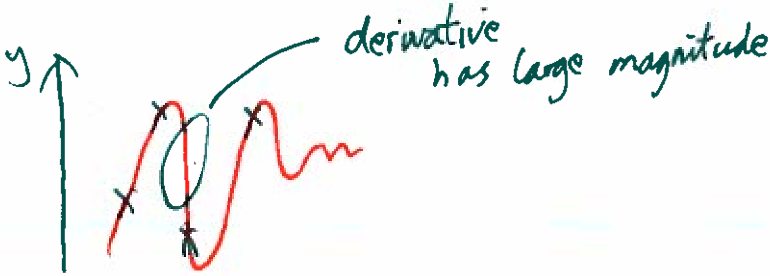


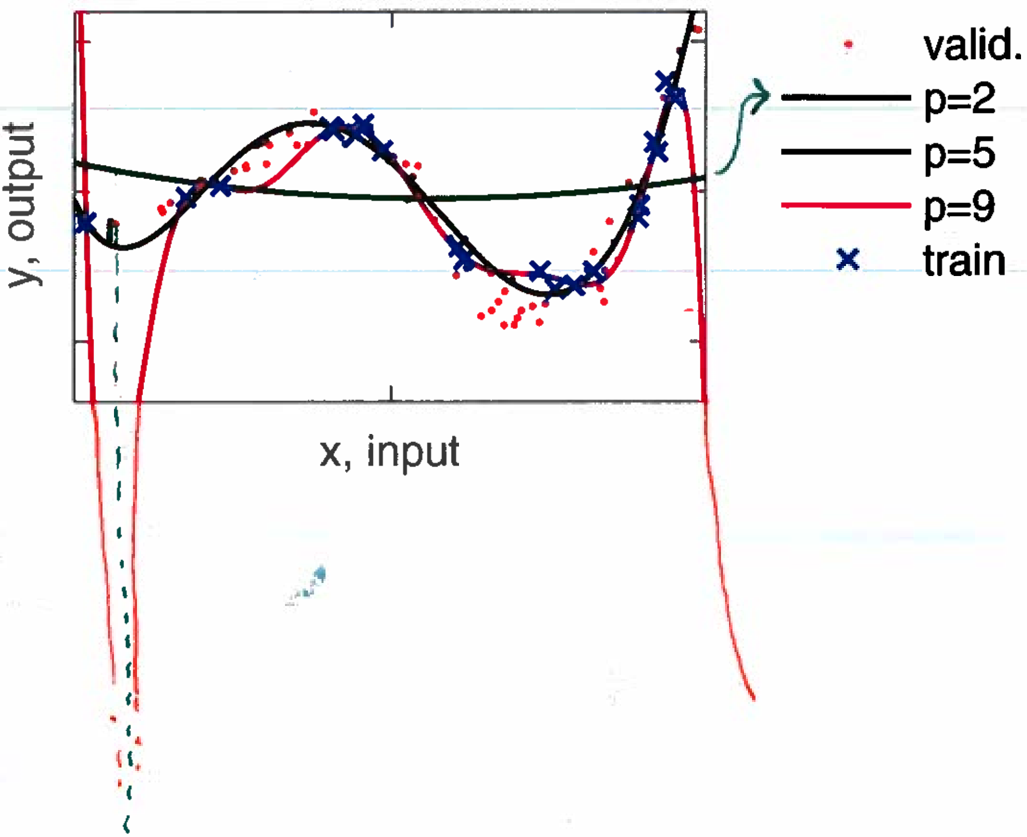
If $\underline{w}$ are bounded then function is bounded



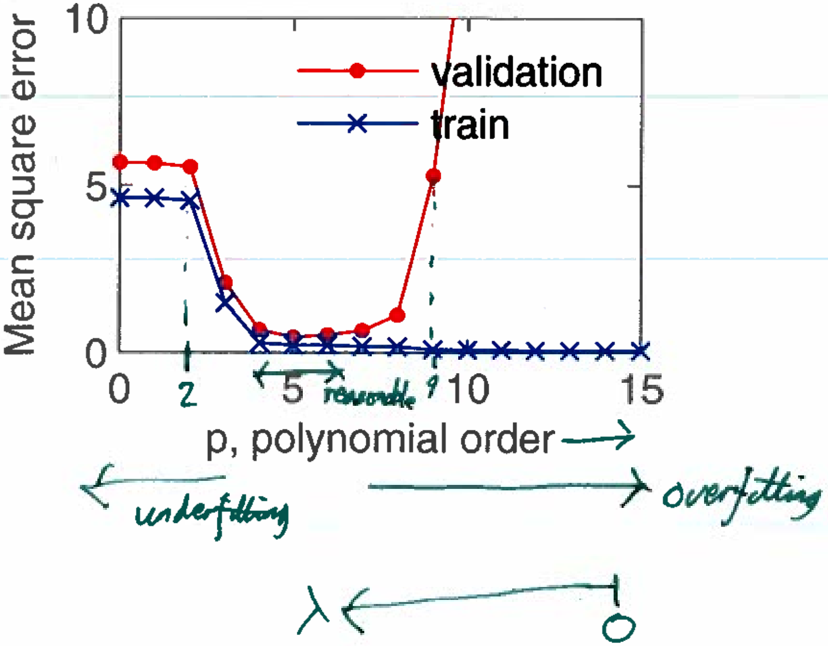
$$f(\underline{x};\underline{w}) = \sum_k w_k \phi_k(\underline{x}) \leqslant \sum_k |w_k| \underbrace{|\phi_k(\underline{x})|}_{\leqslant 1 \text{ often}} \leqslant \sum_k |w_k|$$







2019 L4 5


10
5
0
Mean square error
validation
train
0
2
5
reasonable
9
10
15
p, polynomial order
underfitting
overfitting
λ
0

2019 4 6

Generalization

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

Loss function

Assuming there is some fixed distribution on future inputs & outputs

$dx_1 dx_2 \ldots dx$

$$= \iint L(y, f(\underline{x}))\, p(\underline{x}, y)\, d\underline{x}\, dy$$

$$\text{OR} = \sum_{\underline{x}} \sum_{y} L(y, f(x))\, p(\underline{x}, y)$$

Monte Carlo unbiased approximation

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

$$y^{(m)}, \underline{x}^{(m)} \sim p(\underline{x}, y)$$

Draw samples from a held out test set.

2019 L4 7

Data Splits

Training Set: used to fit $\underline{w}$

Couldn't use (yet): to fit:

- Order polynomial
- # RBFs, # $\sigma(\underline{x})$
- Regularization constant, λ

Validation Set: To "fit λ", model choices
(Dev. set

Test Set: Use to report an estimate of generalization error.

Reading: kaggle blog.