Tutorials:

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

Hypothesis Forum
- Share links, code snippets
- Get code review
- Ask Questions help others
- Post Answers 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 \( f(x) = \mathbf{w}^T \phi(x) \)

Can minimize

\[ \sum_{n} (y^{(n)} - \mathbf{w}^T \phi(x^{(n)}))^2 \]

wrt \( \mathbf{w} \)

\[ = (\mathbf{y} - \Phi \mathbf{w})^T (\mathbf{y} - \Phi \mathbf{w}) \]

\[ \phi(x) = [\phi_1(x) \phi_2(x) ... \phi_k(x)]^T \]

\( \phi_k(x) \) any scalar function
- Monomial, eg. \( x_2, x_3 x_4^3, ... \)
- RBF

\[ \mathbf{w}^T \]

- Sigmoid

\[ \sigma(x) \]
Why are large weights bad?

If \( w \) are bounded, then function is bounded.

\[
f(x; w) = \sum_{k} w_k \phi_k(x)
\leq \sum_{k} |w_k| |\phi_k(x)|
\leq 1 \text{ if } \sum_{k} |w_k| < 1.
\]

Derivative has large magnitude.
Mean square error

\( p, \text{ polynomial order} \)

Validation
Train

Underfitting
Overfitting

2019 4⑥
Generalization

\[ E_{\text{gen}} = \mathbb{E} \left[ L(y, f(x)) \right] \]

\[ p(x, y) \quad \text{Loss function} \]

Assuming there is some fixed distribution on future inputs & outputs \( dx, dy \).

\[ = \int \int L(y, f(x)) p(x, y) \, dx \, dy \]

\[ \text{OR} = \sum_{x} \sum_{y} L(y, f(x)) p(x, y) \]

Monte Carlo approximation

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

\[ y^{(m)}, x^{(m)} \sim p(x, y) \]

Draw samples from a held out test set.

2019 44\( \mathbb{E} \)
Data Splits

Training Set: used to fit w

- Order polynomial
- $\# \text{RBFs}, \ # o(z)$
- Regularization constant: $\lambda$

Couldnt use (yet): to fit:

Validation Set: (Dev. set)

To "fit $\lambda", \ model \ chooses$

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

Reading: Kaggle blog.