Tutorials: • 1st sheet up
  • Meetings next week (TBA soon!)
  • Answers released end next week

Assignment pairs:
  See website

Hypothesis Forum
  - Share links, code snippets
  - Get code review
  - Ask Q's
  - Post answers < help others get feedback
**Linear Regression Reminders**

Model \( f(x) = w^T \phi(x) \)

Can minimize

\[
\sum_n (y^{(n)} - w^T \phi(x^{(n)}))^2 \text{ wrt } w \\
= (y - \Phi w)^T (y - \Phi w)
\]

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

\( \phi_k(x) \) any scalar function

- Monomial, e.g. \( x_2, x_3 x_4^3, \ldots \)

- RBF

\[
\begin{array}{ccc}
\text{\( w^T \)} & \Rightarrow & \text{\( \Phi \)} \\
\text{\( x \)} & & \text{\( \sigma(x) \)}
\end{array}
\]

- Sigmoid
Why are large weights bad?

If $w$ are bounded then function bounded.

Large derivatives are bad.
If $w$ are bounded \( \Rightarrow \) derivatives are bounded.

RBF: Always extrapolates to 0.

\( \sigma(x) \)


- Mean square error vs. polynomial order $p$
- Validation and train error graphs
- Overfitting and underfitting
- Reasonable model selection
- Regularization $\lambda$
Generalization

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

Loss function

Assuming there is some fixed distribution on future inputs & outputs

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

Monte Carlo approximation

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

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

Draw samples from a held out test set.
Data Splits

Training set: fit w

Don't fit:

Order of a polynomial
# of RBFs
Regularization constant $\lambda$

Validation Set

To fit $\lambda$, model choices

Test set

To report estimate of generalization error.

Reading: kaggle blog.