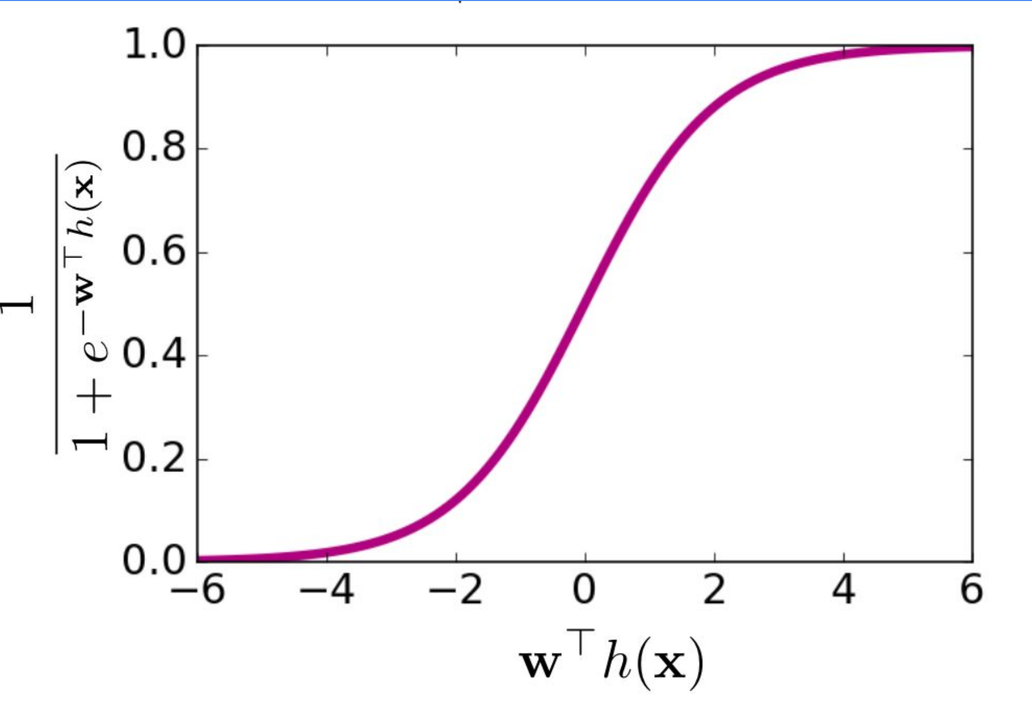
Regularizations in logistic regression

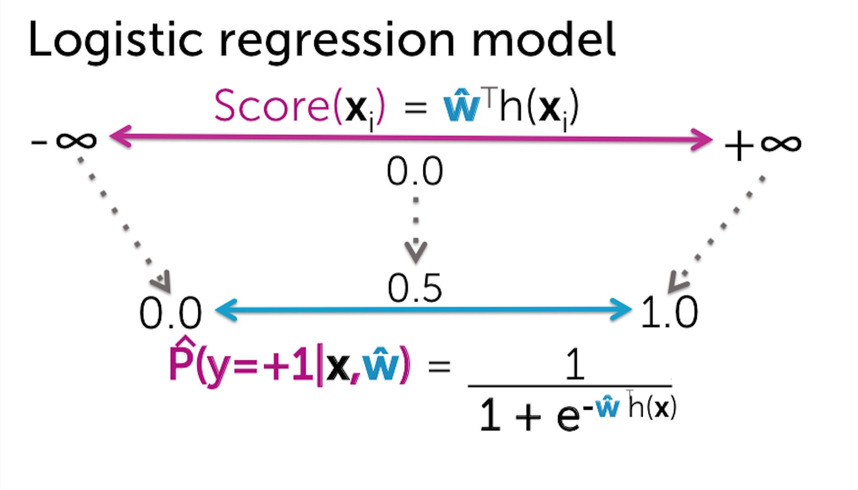
Emma Yu
WiDS ATX
Oct 20, 2016

A special thank you!

Lots of plots and ideas of this talk are taken from excellent notes from Emily Fox @ U Washington

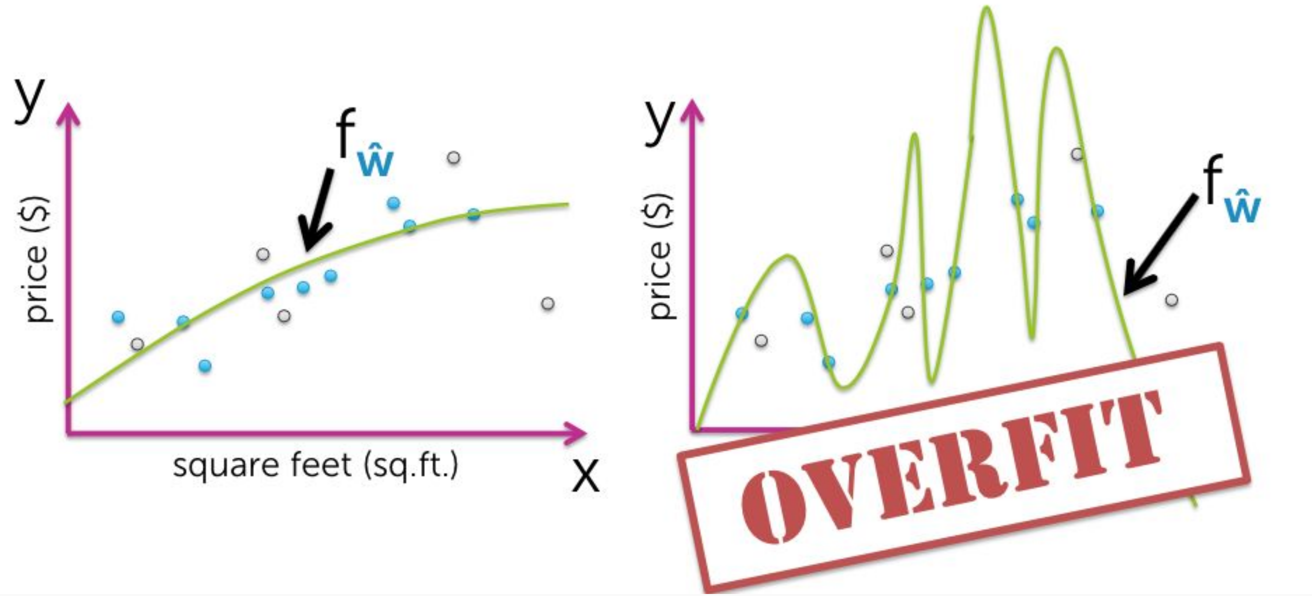
With a logit link, we can apply our knowledge on linear regression to classification problems



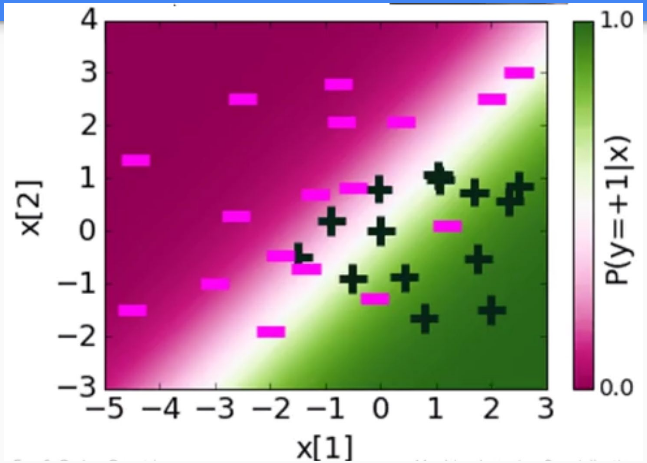


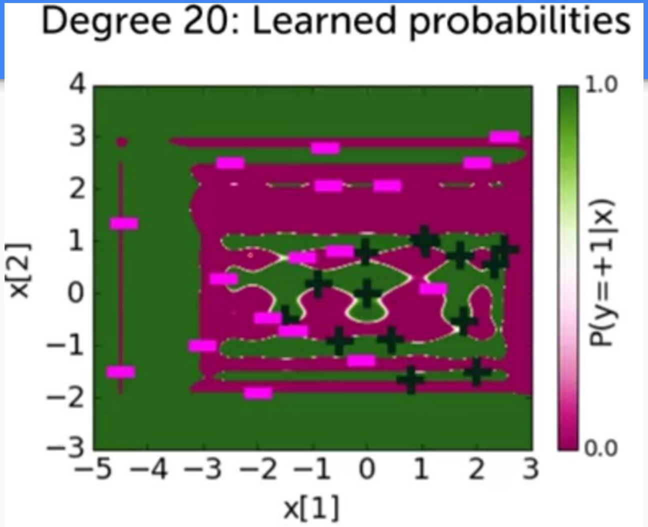
Remember in linear regression we have problem with overfitting?

$$y_i = w_0 + w_1 x_i + w_2 x_i^2 + \dots + w_p x_i^p + \varepsilon_i$$



Signs of overfitting in logistic regression





- Overly complex decision boundary
- Overly large coefficients
- Side effect: “overconfidence” due to large coefficients

Regularizations to the rescue

Instead of maximizing the likelihood function, we maximize

Total quality = quality of fit + measure of model complexity

L1 penalty: $\ell(\mathbf{w}) - \lambda||\mathbf{w}||_1$

L2 penalty: $\ell(\mathbf{w}) - \lambda||\mathbf{w}||_2^2$

(Note that L1 regularization leads to sparsity)

How much to penalize?

Tuning factor lambda controls the model complexity

Optimized through validation/cross-validation

(note that in sklearn, C= 1/lambda)

Demo time!