

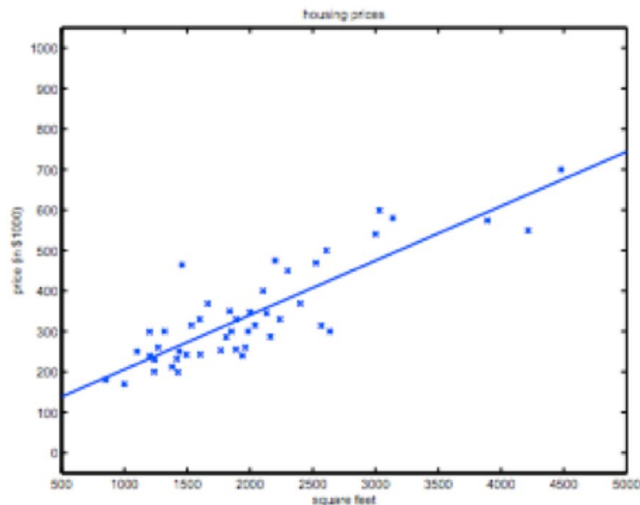
Logistic Regression

GA DAT3

Agenda

1. Linear Regression Review
2. From Linear to Logistic
3. Performance Measures
4. GLMs, Exponential Family
5. Relationship to Naive Bayes

Linear Reg Review



$$h(x) = \sum_{i=0}^n \theta_i x_i = \theta^T x$$

Squared Error Loss

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Why?

Probabilistic Interpretation

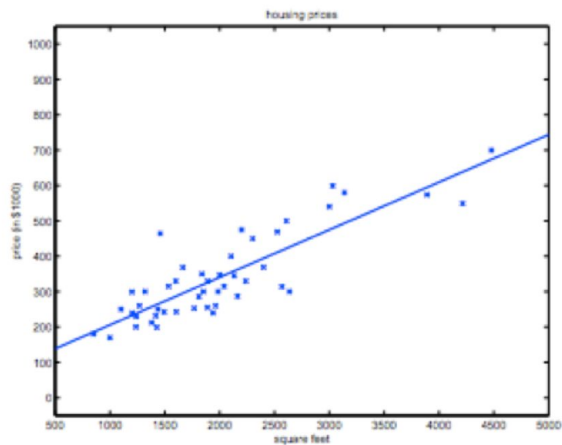
$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

$$p(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right).$$

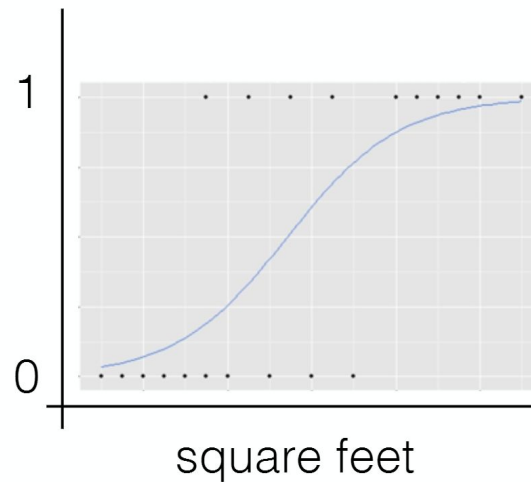
We can minimise **cost**
or
maximise likelihood

What's the likelihood?

Linear to Logistic



bungalow

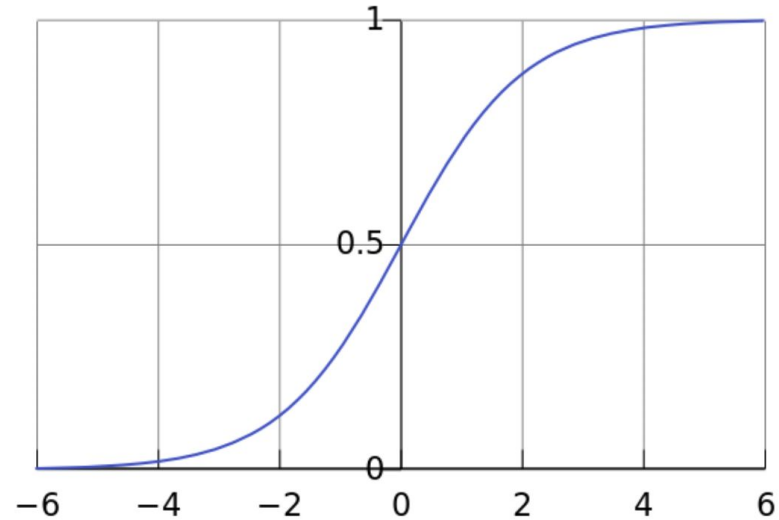


Logistic function

The logistic function always returns a value between zero and one.

$$h(x) = \text{theta_transpose } x$$

$$F(t) = \frac{1}{1 + e^{-t}}$$



Problems with just using linear regression to classify?

Classification vs Clustering?

supervised vs unsupervised

Examples of Classification?

Performance Measures

Predict whether tumors are malignant or benign:

- **Accuracy:** fraction of instances that are classified correctly
 - does not differentiate between malignant tumors that were classified as being benign, and benign tumors that were classified as being malignant.
- In some problems, the costs associated with all types of errors may be the same
- In this problem, failing to identify malignant tumors is likely more severe than failing to identify benign tumors as malignant

Performance Measures

- **True positive:** correctly classifying a malignant tumor
- **True negative:** correctly classifying a benign tumor
- **False positive:** a benign tumor that is incorrectly classifier as being malignant
- **False negative:** a malignant tumor that is incorrectly classifier as being benign

Confusion Matrix

| | Prediction | | |
|--------|------------|-----------|-----------|
| Actual | | 1 | 0 |
| | 1 | TP | FN |
| | 0 | FP | TN |

Performance Measures

- **Accuracy** is the fraction of instances that were classified correctly

$$\text{ACC} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

Performance Measures

- ***Precision*** is the fraction of the tumors that were predicted to be malignant that are actually malignant.

$$P = TP / (TP + FP)$$

Performance Measures

- ***Recall*** (or True Positive Rate) is the fraction of malignant tumors that the system identified.

$$R = TP / (TP + FN)$$

Performance Measures

- ***Recall*** (or True Positive Rate) is the fraction of malignant tumors that the system identified.

$$R = TP / (TP + FN)$$

Performance Measures

- ***Fall-out*** or *false positive rate* (***FPR***):

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

Generalised Linear Models (GLMs)

We've seen

$y \mid x \sim N(\mu, \sigma^2)$ \longrightarrow linear regression

$y \mid x \sim \text{Bernoulli}(\phi)$ \longrightarrow logistic classification

Can we find common ground?

The Exponential Family

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

The Exponential Family: Bernoulli

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

$$\begin{aligned} p(y; \phi) &= \phi^y (1 - \phi)^{1-y} \\ &= \exp(y \log \phi + (1 - y) \log(1 - \phi)) \\ &= \exp\left(\left(\log\left(\frac{\phi}{1 - \phi}\right)\right) y + \log(1 - \phi)\right) \end{aligned}$$

What are...

$$\eta =$$

$$T(y) =$$

$$a(\eta) =$$

$$=$$

$$b(y) =$$

The Exponential Family: Bernoulli

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

$$\begin{aligned} p(y; \phi) &= \phi^y (1 - \phi)^{1-y} \\ &= \exp(y \log \phi + (1 - y) \log(1 - \phi)) \\ &= \exp \left(\left(\log \left(\frac{\phi}{1 - \phi} \right) \right) y + \log(1 - \phi) \right) \end{aligned}$$

What are...

$$\eta = \log(\phi / (1 - \phi)).$$

$$T(y) = y$$

$$\begin{aligned} a(\eta) &= -\log(1 - \phi) \\ &= \log(1 + e^\eta) \end{aligned}$$

$$b(y) = 1$$

The Exponential Family: Normal

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

$$p(y; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \mu)^2\right)$$

What are...

$$\eta =$$

$$T(y) =$$

$$a(\eta) =$$

$$=$$

$$b(y) =$$

The Exponential Family: Normal

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

$$\begin{aligned} p(y; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \cdot \exp\left(\mu y - \frac{1}{2}\mu^2\right) \end{aligned}$$

What are...

$$\begin{aligned} \eta &= \mu \\ T(y) &= y \\ a(\eta) &= \mu^2/2 \\ &= \eta^2/2 \\ b(y) &= (1/\sqrt{2\pi}) \exp(-y^2/2) \end{aligned}$$

Okay, okay...so who cares?

Constructing GLMs

1. Assume $y \mid x; \theta \sim \text{ExponentialFamily}(\eta)$
2. Given x , we want to predict $T(y)$, usually $= y$.
We choose $h(x) = E[y|x]$
3. Further assume $\eta = \theta^T x$

So we have a machinery we can crank

Constructing GLMs

Linear Regression

$$\begin{aligned}h_{\theta}(x) &= E[y|x; \theta] \\&= \mu \\&= \eta \\&= \theta^T x.\end{aligned}$$

Logistic Classification

$$\begin{aligned}h_{\theta}(x) &= E[y|x; \theta] \\&= \phi \\&= 1/(1 + e^{-\eta}) \\&= 1/(1 + e^{-\theta^T x})\end{aligned}$$

returns more than
2 labels



Coincidentally, this is how we get softmax regression...

Relationship to Naïve Bayes

Assuming $y \mid x \sim \text{some distribution}$

Assuming $x \mid y \sim \text{some distribution}$

e.g.

text classification

Gaussian Discriminant Analysis (GDA)

Q??
