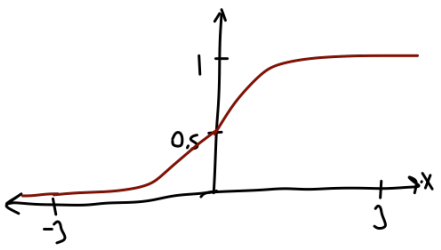
*Linear Regression is not suitable for classification

Binary classification is when there are 2 options.

Sigmoid Function (Logistic Function)



outputs between 0 and 1

$g(x) = \frac{1}{1+e^{-x}}$

Logistic Regression

$z = \vec{w} \cdot \vec{x} + b \quad g(z) = \frac{1}{1+e^{-z}}$

* Outputs probability of y being 1

$f_{\vec{w},b}(\vec{x}) = P(y=1 \mid \vec{x}; \vec{w}, b)$

*In logistic regression, decision boundary can be made more complex (curved) with using higher rank polynomials

Cost Function For Logistic function

Using squared error cost function for logistic regression makes cost function non-convex meaning contains too many local minima, so it is not suitable.

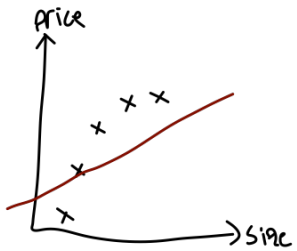
Let J to be $\frac{1}{m}\sum_{i=1}^{m} L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})$ and $L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})$ to be $\begin{cases} -\log f_{\vec{w},b}(\vec{x}^{(i)}), & y^{(i)}=1 \\ -\log[1-f_{\vec{w},b}(\vec{x}^{(i)})], & y^{(i)}=0 \end{cases}$

L can be written in the form: $-y^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)})) - (1-y^{(i)}) \log(1-f_{\vec{w},b}(\vec{x}^{(i)}))$

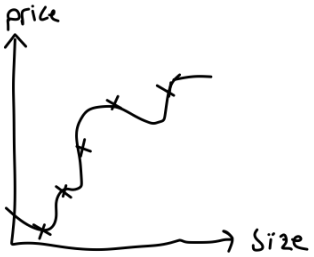
and J becomes: $-\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)}) + (1-y^{(i)}) \log(1-f_{\vec{w},b}(\vec{x}^{(i)})]$

Underfitting: If model we use doesn't have the capability to fit the data, it called underfitting (High bias)

Overfitting: If model fits much more than needed to training set, and doesn't generalize well, then it is called overfitting. (High variance)



Underfitting



Overfitting

* Regularization keeps all of the features, but prevents the features from having an overly large effect, which is what sometimes can cause overfitting.

Adressing Overfitting

Options:

1. Collect more data
2. Feature selection
3. Regularization

$$J(\vec{w},b)=\frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)})-y^{(i)}\right)^2 + \underbrace{\frac{\lambda}{2m}\sum_{j=1}^{n} w_n^2}_{\text{Regularization term}} + \frac{\lambda}{2m}b^2$$

(can include or exclude b, has a little effect)

λ: Regularization parameter

∇ Choose λ carefully!!