# Lecture 3

Sunday, April 19, 2020 11:56 AM

## Locally Weighted & Logistic Regression

Locally weighted regression: Fit $\theta$ to minimize $\sum_{i=1}^{m} w^{(i)}(y^{(i)} - \theta^T x^{(i)})^2$

where $w^{(i)}$ is a "weighting" function

$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$

note that $x^{(i)}$ is $i$th training example.

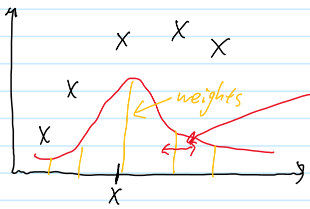
$x$ is the location where you want to make a prediction.

$w^{(i)}$ is the function outputting a value between 0 and 1 which tells you how much you should pay attention to $(x^{(i)}, y^{(i)})$ when fitting.

if $|x^{(i)} - x|$ is small, $w^{(i)} \approx 1$

if $|x^{(i)} - x|$ is large, $w^{(i)} \approx 0$

The weighting function $w^{(i)}$ effect is to sum over all examples that have square error that is small (in other word, it helps to eliminate the examples where the square error is large).




What range of neighbor should you take into consideration?
$\tau$: "bandwidth"

When to use locally weighted regression?
when you have relatively low dimension (i.e. when $n$ is small)

---

## Why use least squares?

using house pricing example:

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

"error" unmodeled effects, random noises

$\varepsilon^{(i)} \sim N(0, \sigma^2)$ Normal distribution with mean 0, variance $\sigma^2$

assume iid (independently indentically distributed) error term for one house is independent from other houses.

$$P(\varepsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(\varepsilon^{(i)})^2}{2\sigma^2}\right)$$

implies that: $$p(y^{(i)} \mid x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$