Support Vector Machines

- Alternative view of logistic regression

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

If $y = 1$, we want $h_\theta(x) \approx 1, \theta^T x \gg 0$

If $y = 0$, we want $h_\theta(x) \approx 0, \theta^T x \ll 0$

- Cost function

Logistic regression:

$$\min \frac{1}{m}\left(\sum_{i=1}^{m} y^{(i)} \left(-log h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\left(-\log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right)\right) + \frac{\lambda}{2m}\sum_{i=1}^{n} \theta_j^2$$

Support vector machine:

$$\min C \sum_{i=1}^{m} \left(y^{(i)} cost_1\left(\theta^T x^{(i)}\right) + \left(1 - y^{(i)}\right) cost_0\left(\theta^T x^{(i)}\right)\right) + \frac{1}{2}\sum_{i=1}^{n} \theta_j^2$$

- Kernel

Given x, compute new feature depending on proximity to landmarks $l^{(1)}, l^{(2)}, l^{(3)}$

$$f_1 = similarity\left(x, l^{(1)}\right) = \exp\left(-\frac{\left\|x - l^{(1)}\right\|^2}{2\sigma^2}\right)$$

$$f_2 = similarity\left(x, l^{(2)}\right) = \exp\left(-\frac{\left\|x - l^{(2)}\right\|^2}{2\sigma^2}\right)$$

$$f_3 = similarity(x, l^{(3)}) = \exp(-\frac{\left\|x - l^{(3)}\right\|^2}{2\sigma^2})$$

Kernels and Similarity

$$f_1 = similarity(x, l^{(1)}) = \exp(-\frac{\left\|x - l^{(1)}\right\|^2}{2\sigma^2})$$

If $x \approx l^{(1)}$:

$$f_1 \approx \exp(-\frac{0^2}{2\sigma^2}) \approx 1$$

If x is far from $l^{(1)}$:

$$f_1 = \exp(-\frac{(large\ number)^2}{2\sigma^2}) \approx 0$$

- SVM with Kernels

Hypothesis: Given x, compute features $f \in R^{m+1}$

Predict “y=1” if $\theta^T f \geq 0$

Training:

$$\min C \sum_{i=1}^{m} \left(y^{(i)} cost_1(\theta^T f^{(i)}) + (1 - y^{(i)}) cost_0(\theta^T f^{(i)})\right) + \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$