Support Vector Machine

SVM Decision Boundary

- Logistic Regression: A large C parameter tells the svm to try to classify all the example correctly.

$$\min_\theta C \cdot \sum_{i=1}^{m} \left[y^{(i)} \cdot \text{cost}_1(\theta^T \cdot x^{(i)}) + (1-y^{(i)}) \cdot \text{cost}_0(\theta^T \cdot x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\Rightarrow \min_\theta \frac{1}{2} \sum_{j=1}^{n} \theta_j^2, \text{ such that whenever } \begin{cases} y^{(i)} = 1, \ \theta^T \cdot x^{(i)} \geq 1 \\ y^{(i)} = 0, \ \theta^T \cdot x^{(i)} \leq -1 \end{cases}$$

$C = \frac{1}{\lambda}$

⇓ large C, high variance; small C, high bias.

Mathematics Behind SVM:

$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$.

$\|u\|$ = length of vector $u = \sqrt{u_1^2 + u_2^2} \in \mathbb{R}$

- Use SVM software package (e.g.: liblinear, libsvm, ...) to solve for parameters θ.

Need to specify:

i). choice of parameter C.

ii). choice of kernel (similarity function).

Gaussian kernel: $f_i = \exp\left(-\frac{\|x - l^{(i)}\|^2}{2\sigma^2}\right)$, where $l^{(i)} = x^{(i)}$.

need to choose σ^2

$$K_{\text{gaussian}}(x^{(i)}, x^{(j)}) = \exp\left(-\frac{\|x^{(i)} - x^{(j)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{k=1}^{n} (x_k^{(i)} - x_k^{(j)})^2}{2\sigma^2}\right)$$

Polynomial Kernel

n = number of features, m = number of training examples.

i). if n is large, use logistic regression or SVM without a kernel

ii). if n is small, m is intermediate, use SVM with Gaussian kernel

iii). If n is small, m is large, create/add more features, then do i).

Clustering.

K-Mean algorithm:

1. Randomly initialize K cluster centroids $u_1, u_2, \cdots, u_k \in R^n$.
2. Repeat {

{ for $i=1$ to m (m is the # of examples)

$c^{(i)}$ = index (from 1 to K) of cluster centroid closest to $x^{(i)}$.

} ← cluster assignment step.

for $k=1$ to K:

u_k = average of points assigned to cluster k

}.

Optimization objective:

Assigned cluster centroid.

$$\min J(c^{(1)}, \cdots, c^{(m)}, u_1, \cdots, u_k) = \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - u_{c^{(i)}} \|^2$$

choose the number of K:

Elbow method: