Classifier: SVM vs Logistic Regression

Learning Type: Supervised, Task: Classification, Algorithm: SVM and Logistic Regression

Model $h(x)$:
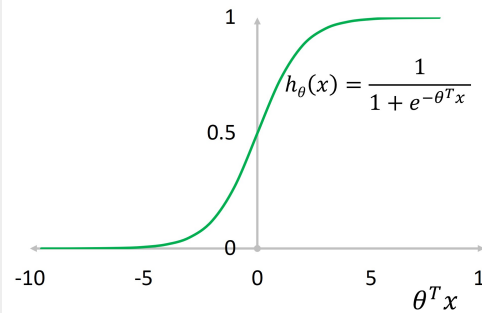
Logistic Regression (sigmoid):
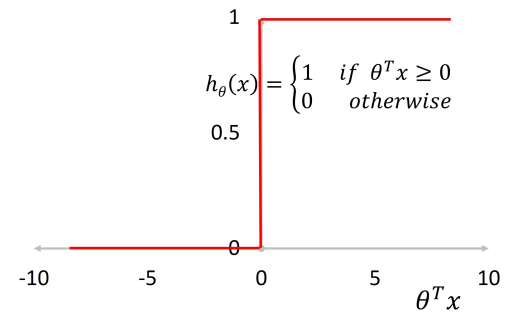$h_\theta(x)_{LR} = \frac{1}{1+e^{-\theta^T x}} = \frac{e^{\theta^T x}}{e^{\theta^T x}+1}$

Support Vector Machine:
$h_\theta(x)_{SVM} = \begin{cases} 1 & \text{if } \theta^T x >= 0 \\ 0 & \text{otherwise} \end{cases}$

Logistic Regression h(x)

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

Support Vector Machine h(x)

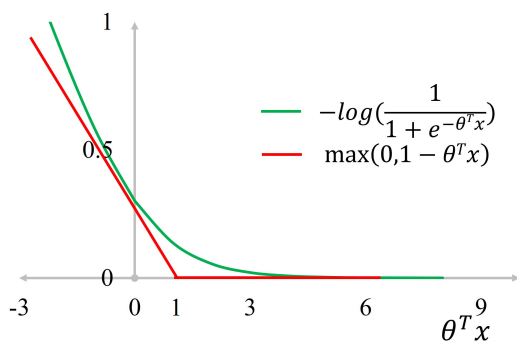$h_\theta(x) = \begin{cases} 1 & \text{if } \theta^T x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Evaluation Criteria: Cost Function

Cross Entropy or Log Loss for LR: $J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right)\log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$

Hinge Loss for SVM: $J(\theta) = C[\sum_{i=1}^{m} y^{(i)} \max\left(0, 1 - \theta^T x\right) + \left(1 - y^{(i)}\right)\max\left(0, 1 + \theta^T x\right)] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$

where m = no. of examples; n = no. of features; X = dataset and θ = model weights

Loss for truth class $y = 1$

— $-log(\frac{1}{1 + e^{-\theta^T x}})$
— $\max(0, 1 - \theta^T x)$

Loss for truth class $y = 0$

— $-log(1 - \frac{1}{1 + e^{-\theta^T x}})$
— $\max(0, 1 + \theta^T x)$
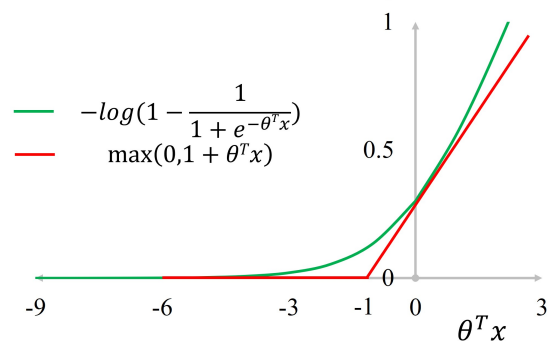
More Details about SVM and Logistic Regression

	Logistic Regression	**Support Vector Machine**
Regularization	$\lambda \uparrow$ Overfitting $\downarrow$	C $\uparrow$ Overfitting $\uparrow$ C $= \frac{1}{\lambda}$
MultiClass	One-vs-all	One-vs-all
NonLinear Dataset	Nonlinear transformation of $\theta^T x$	Kernel function to convert X to higher dimension and use linear hyperplane
Class Margin	Effect with λ	C $\uparrow$ Loss $\uparrow$ Margin $\downarrow$ Hard SVM C $\downarrow$ Loss $\downarrow$ Margin $\uparrow$ Soft SVM

Underfit ($\lambda\uparrow$)

Overfit ($\lambda\downarrow$)

Soft SVM (C$\downarrow$)

Hard SVM (C$\uparrow$)