CHEAT SHEET

Loss Function

Loss		Usage	Comments
Absolute	$\left\|h_{\mathbf{w}}\left(\mathbf{x}_i\right)-y_i\right\|$	Applicable to all regression problems	This function mitigates outliers by only taking the absolute difference. However, it is not differentiable at 0, the value we are trying to optimize to. For noisy data, it will try to predict the median.
Exponential	$e^{-h_w(x_i)y_i}$	AdaBoost	This function is very aggressive. The loss of a mis-prediction increases exponentially with the value of $-h_w(x_i)y_i$ .This can lead to nice convergence results, for example in the case of Adaboost, but it can also cause problems with noisy data.
Hinge	$\max\left[1-h_{\mathbf{w}}\left(\mathbf{x}_i\right)y_i,0\right]^p$	• Standard SVM $(p=1)$ • Differentiable Squared Hingeless SVM $(p=2)$	When used for Standard SVM, the loss function denotes the size of the margin between linear separator and its closest points in either class. Only differentiable everywhere with $(p=2)$.
Log	$\log\left(1+e^{-h_{\mathbf{w}}(\mathbf{x}_i)y_i}\right)$	Logistic Regression	One of the most popular loss functions in Machine Learning, since its outputs are well-calibrated probabilities.
Squared	$\left(h_{\mathbf{w}}\left(\mathbf{x}_i\right)-y_i\right)^2$	Applicable to all regression problems	This function is differentiable everywhere and easy to optimize; it has a closed-form solution. Given noisy data, it will try to predict the mean. However, because the difference is squared, outliers can easily dominate training as their large differences get even larger.
Zero-One	$\mathbf{1}_{\operatorname{sign}(h_{\mathbf{w}}(\mathbf{x}_i))\neq y_i}$	Actual Classification Loss	Non-continuous and thus impractical to optimize.