eCornell

CHEAT SHEET

Famous Regularizers

Loss Type		Comments
Ordinary Least Squres	$\min_{\mathbf{w}} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^\top \mathbf{x}_i - y_i)^2$	• Squared loss • No regularization • Closed form solution: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}y^\top$ $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$ $y = [y_1, ..., y_n]$
Ridge Regression	$\min_{\mathbf{w}} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^\top \mathbf{x}_i - y_i)^2 + \lambda\|\|\mathbf{w}\|\|_2^2$	• Squared loss • l_2-regularization • Closed form solution: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top + \lambda\mathbb{I})^{-1}\mathbf{X}y^\top$
Lasso	$\min_{\mathbf{w}} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^\top \mathbf{x}_i - y_i)^2 + \lambda\|\|\mathbf{w}\|\|_1$	• Also known as l_1-regularization • + Sparsity inducing, helps feature selection • + Convex • - Not strictly convex (no unique solution) • - Not differentiable (at 0) • Solve with (sub)-gradient descent or SVEN
Elastic Net	$\min_{\mathbf{w}} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^\top \mathbf{x}_i - y_i)^2 + a\|\|\mathbf{w}\|\|_1 + (1-a)\|\|\mathbf{w}\|\|_2^2$	• + Strictly convex (i.e. unique solution) • + Sparsity incuding (good for feature selection) • Disadvantage: Non-differentiable
Logistic Regression	$\min_{\mathbf{w},b} \frac{1}{n}\sum_{i=1}^{n}\log(1+e^{-y(\mathbf{w}^\top \mathbf{x}+b)})$	• Often l_1 or l_2 regularized • Solve with gradient descent. • Calibrated output probabilities: $P(y\|x) = \frac{1}{1+e^{-y(\mathbf{w}^\top x+b)}}$
Linear Support Vector Machine	$\min_{\mathbf{w},b} C\sum_{i=1}^{n}\max[1-y_i(\mathbf{w}^\top \mathbf{w}_i + b), 0] + \|\|\mathbf{w}\|\|_2^2$	• Typically l_2 regularized (sometimes l_1) • When kernelized leads to sparse solutions • Kernelized version can be solved very efficiently with specialized algorithms