Pset 3: SVMs

- hinge loss $= \max(0, 1 - y_i \cdot (x_i \cdot w))$

true label

-1 or +1

ith row of input

weight vector

score = data · weight vector for ith

margin

regularization term λ penalizes large weights

- $$w^* = \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \frac{\lambda}{2} \|w\|^2 + \frac{1}{N} \sum_{i=1}^{N} [1 - y_i (w \cdot x_i)]_+$$

where $(x_1, y_1) \ldots (x_N, y_N) \in \mathbb{R}^d \times \{\pm 1\}$ $[\cdot]_+ = \max(0, \cdot)$

$\frac{\partial}{\partial w} \max(0, \text{margin}) = -y \times X$

- gradient $\begin{cases} -y \times X & \text{if } 1 - y \times \text{score} > 0 \\ 0 & \text{" } \leq 0 \end{cases}$

misclassified or within margin boundary

correctly classified and outside margin boundary (not support vec

function is flat for margins ≤ 0

" " $+ \lambda \times w$ if " "

" " $+ \lambda \times w$ if " "

when you add regularization

Gaussian / RBF kernel:

$*\exp(x) = e^x$

- $K(x, y) = \exp(-\gamma \|x - y\|^2)$

Euclidean distance between each pair of points.

kernel SVM maintains K support vectors and learning weights $\alpha_1 \ldots \alpha_K \geq 0$

positive constant $\gamma = \frac{1}{2\sigma^2}$

variance

$X = N \overset{d}{[\quad]}$

$Y = M \overset{d}{[\quad]}$

$K = N \overset{M}{[\quad]}$

each entry i, j contains kernel function result using $X[i]$ and $Y[j]$

number of support vectors or can be # of data points to compare with X matrix

- learning rate is the amount the weights are updated between 0 and 1, also called step size
- pegasos algorithm kernalizes primal objective with