Chapter 56

ν-SV Regression

56.1 ν-SV Regression; Derivation of the Dual

Let $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ be a set of observed data usually called a set of *training data*, with $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. As in Chapter 55, we form the $m \times n$ matrix X where the row vectors $x_i^\top$ are the *rows* of X. Our goal is to learn an affine function f of the form $f(x) = x^\top w + b$ that fits the set of training data, but does not penalize errors below some given $\epsilon \geq 0$. Geometrically, we view the pairs (x_i, y_i) are points in $\mathbb{R}^{n+1}$, and we try to fit a hyperplane $H_{w,b}$ of equation

$$(w^\top \;\; -1) \begin{pmatrix} x \\ z \end{pmatrix} + b = w^\top x - z + b = 0$$

that best fits the set of points (x_i, y_i) (where $(x, z) \in \mathbb{R}^{n+1}$). We seek an $\epsilon > 0$ such that most points (x_i, y_i) are inside the slab (or tube) of width 2ϵ bounded by the hyperplane $H_{w,b-\epsilon}$ of equation

$$(w^\top \;\; -1) \begin{pmatrix} x \\ z \end{pmatrix} + b - \epsilon = w^\top x - z + b - \epsilon = 0$$

and the hyperplane $H_{w,b+\epsilon}$ of equation

$$(w^\top \;\; -1) \begin{pmatrix} x \\ z \end{pmatrix} + b + \epsilon = w^\top x - z + b + \epsilon = 0.$$

Observe that the hyperplanes $H_{w,b-\epsilon}, H_{w,b}$ and $H_{w,b+\epsilon}$ intersect the z-axis when $x = 0$ for the values $(b - \epsilon, b, b + \epsilon)$. Since $\epsilon \geq 0$, the hyperplane $H_{w,b-\epsilon}$ is below the hyperplane $H_{w,b}$ which is below the hyperplane $H_{w,b+\epsilon}$. We refer to the lower hyperplane $H_{w,b-\epsilon}$ as the *blue margin*, to the upper hyperplane $H_{w,b+\epsilon}$ as the *red margin*, and to the hyperplane $H_{w,b}$ as the *best fit hyperplane*. Also note that since the term $-z$ appears in the equations of these hyperplanes, points for which $w^\top x - z + b \leq 0$ are *above* the hyperplane $H_{w,b}$, and points for which $w^\top x - z + b \geq 0$ are *below* the hyperplane $H_{w,b}$ (and similarly for $H_{w,b-\epsilon}$ and