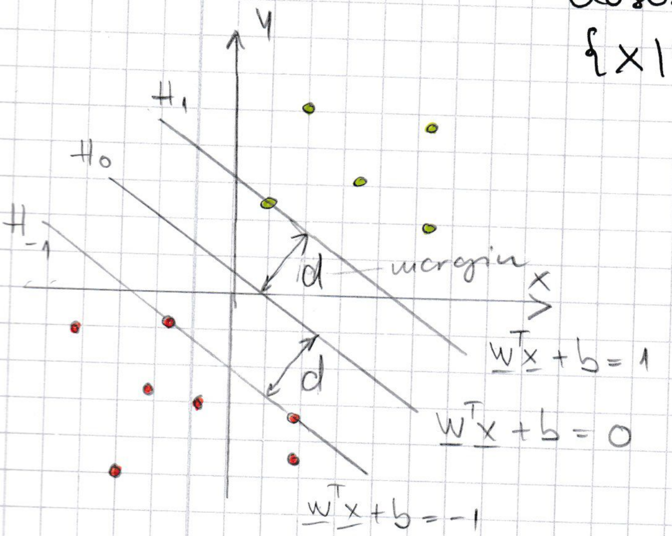
H. 9.1. a)

$$\min_{\alpha=1,\dots,p} |\underline{w}^T \underline{x}^{(\alpha)} + b| \overset{!}{=} 1$$

$d(\underline{x}^{(\alpha)}, \underline{w}, b)$ - Euclidian distance of sample $\underline{x}^{(\alpha)}$ to the closest point on the decision boundary $\{x | y = 0\}$



Distance from point (x_0, y_0) to line $Ax + By + C = 0$ is:

$$\frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

$$\Rightarrow d = \frac{|\underline{w}^T \underline{x}^{(\alpha)} + b|}{\sqrt{\underline{w}^2}} = \frac{1}{\|\underline{w}\|}$$

The min distance of point $\underline{x}^{(\alpha)} \in \mathbb{R}^d$ to the decision boundary

$$\Rightarrow \quad d(\underline{x}^{(\alpha)}, \underline{w}, b) \geq \frac{1}{\|\underline{w}\|}$$

H. 9.1. b)

C-SVM defines slack variables $\varphi_\alpha \geq 0$ for all samples and the primal problem is

$$\min_{\underline{w}, b} \frac{1}{2} \|\underline{w}\|^2 + \frac{C}{P} \sum_{\alpha=1}^{P} \varphi_\alpha$$

subject to: $$y_T^{(\alpha)} (\underline{w}^T \underline{x}^{(\alpha)} + b) \geq 1 - \varphi_\alpha$$

and $$\varphi_\alpha \geq 0 \quad , \forall \alpha$$

the Lagrangian of the primal optimization problem of the C-SVM:

$$\mathcal{L}_P = \frac{1}{2} \|\underline{w}\|^2 + \frac{C}{P} \sum_{\alpha=1}^{P} \varphi_\alpha - \sum_{\alpha=1}^{P} \lambda_\alpha \left[y_T^{(\alpha)} (\underline{w}^T \underline{x}^{(\alpha)} + b) - (1 - \varphi_\alpha) \right] - \sum_{\alpha=1}^{P} \mu_\alpha \varphi_\varphi$$

which we minimize with regard to w, b and φ_α

$$\frac{\partial \mathcal{L}_P}{\partial \underline{w}} = \underline{w} - \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} \underline{x}^{(\alpha)} = 0 \quad \Rightarrow \quad \underline{w} = \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} \underline{x}^{(\alpha)}$$

$$\frac{\partial \mathcal{L}_P}{\partial b} = - \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} = 0 \quad \Rightarrow \quad \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} = 0 \qquad (1)$$

$$\frac{\partial \mathcal{L}_P}{\partial \varphi_\alpha} = \frac{C}{P} - \lambda_\alpha - \mu_\alpha = 0 \quad \Rightarrow \quad \mu_\alpha = \frac{C}{P} - \lambda_\alpha \qquad (2)$$

subject to: $\lambda_\alpha, \mu_\alpha, \varphi_\alpha \geq 0 \quad , \forall \alpha$ (3)

Lagrange dual:

$$\mathcal{L}_D = \frac{1}{2} \sum_{\alpha,\beta=1}^{P} \lambda_\alpha \lambda_\beta y_T^{(\alpha)} y_T^{(\beta)} (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \frac{C}{P} \sum_{\alpha=1}^{P} \varphi_\alpha - \sum_{\alpha=1} \lambda_\alpha y_T^{(\alpha)} \ldots$$

on the next page

Lagrange dual:

We get the expression for the Lagrange dual when we plug (1), (2) and (3) into the expression for the Lagrange primal.

$$\mathcal{L}_D = \frac{1}{2}\sum_{\alpha,\beta=1}^{P} \lambda_\alpha \lambda_\beta y_T^{(\alpha)} y_T^{(\beta)} \left(\underline{x}^{(\alpha)}\right)^T \left(\underline{x}^{(\beta)}\right) + \frac{C}{P}\sum_{\alpha=1}^{P} \varphi_\alpha - \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} \left(\underline{x}^{(\alpha)}\right)^T \sum_{\beta=1}^{P} \lambda_\beta y_T^{(\beta)} \underline{x}^{(\beta)}$$

$$\cancelto{0}{- \sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} b} + \sum_{\alpha=1}^{P} \lambda_\alpha - \sum_{\alpha=1}^{P} \lambda_\alpha \varphi_\alpha - \sum_{\alpha=1}^{P} \left(\frac{C}{P} - \lambda_\alpha\right) \varphi_\alpha$$

$$\mathcal{L}_D = \frac{1}{2}\sum_{\alpha,\beta=1}^{P} \lambda_\alpha \lambda_\beta y_T^{(\alpha)} y_T^{(\beta)} \left(\underline{x}^{(\alpha)}\right)^T \underline{x}^{(\beta)} + \cancel{\frac{C}{P}\sum_{\alpha=1}^{P} \varphi_\alpha} - \sum_{\alpha,\beta=1}^{P} \lambda_\alpha \lambda_\beta y_T^{(\alpha)} y_T^{(\beta)} \left(\underline{x}^{(\alpha)}\right)^T \underline{x}^{(\beta)}$$

$$+ \sum_{\alpha=1}^{P} \lambda_\alpha - \cancel{\sum_{\alpha=1}^{P} \lambda_\alpha \varphi_\alpha} - \cancel{\frac{C}{P}\sum_{\alpha=1}^{P} \varphi_\alpha} + \cancel{\sum_{\alpha=1}^{P} \lambda_\alpha \varphi_\alpha}$$

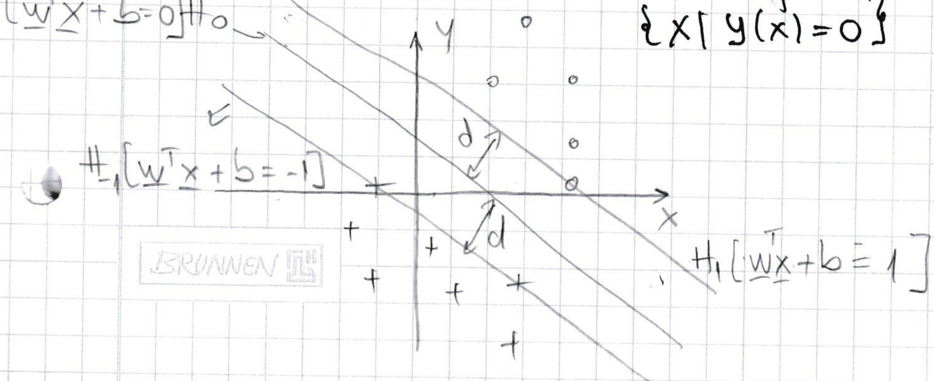
$$\mathcal{L}_D = -\frac{1}{2}\sum_{\alpha=1}^{P}\sum_{\beta=1}^{P} \lambda_\alpha \lambda_\beta y_T^{(\alpha)} y_T^{(\beta)} \left(\underline{x}^{(\alpha)}\right)^T \underline{x}^{(\beta)} + \sum_{\alpha=1}^{P} \lambda_\alpha$$

with $0 \leq \lambda_\alpha \leq \frac{C}{P}, \forall \alpha$ and $\sum_{\alpha=1}^{P} \lambda_\alpha y_T^{(\alpha)} = 0$

a)

$$\min_{\alpha=1,\dots,P} \left| \underline{w}^T \underline{x}^{(\alpha)} + b \right| \stackrel{!}{=} 1$$

$d(\underline{x}^{(\alpha)}, \underline{w}, b)$ - Euclidian distance of sample $\underline{x}^{(\alpha)}$ to the closest point on the decision boundary $\{x \mid y(x) = 0\}$



distance from point (x_0, y_0) to line $Ax + By + C = 0$ is calculated as:

$$\frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$