SOFT MARGIN SVM

Date: / /

Page No.

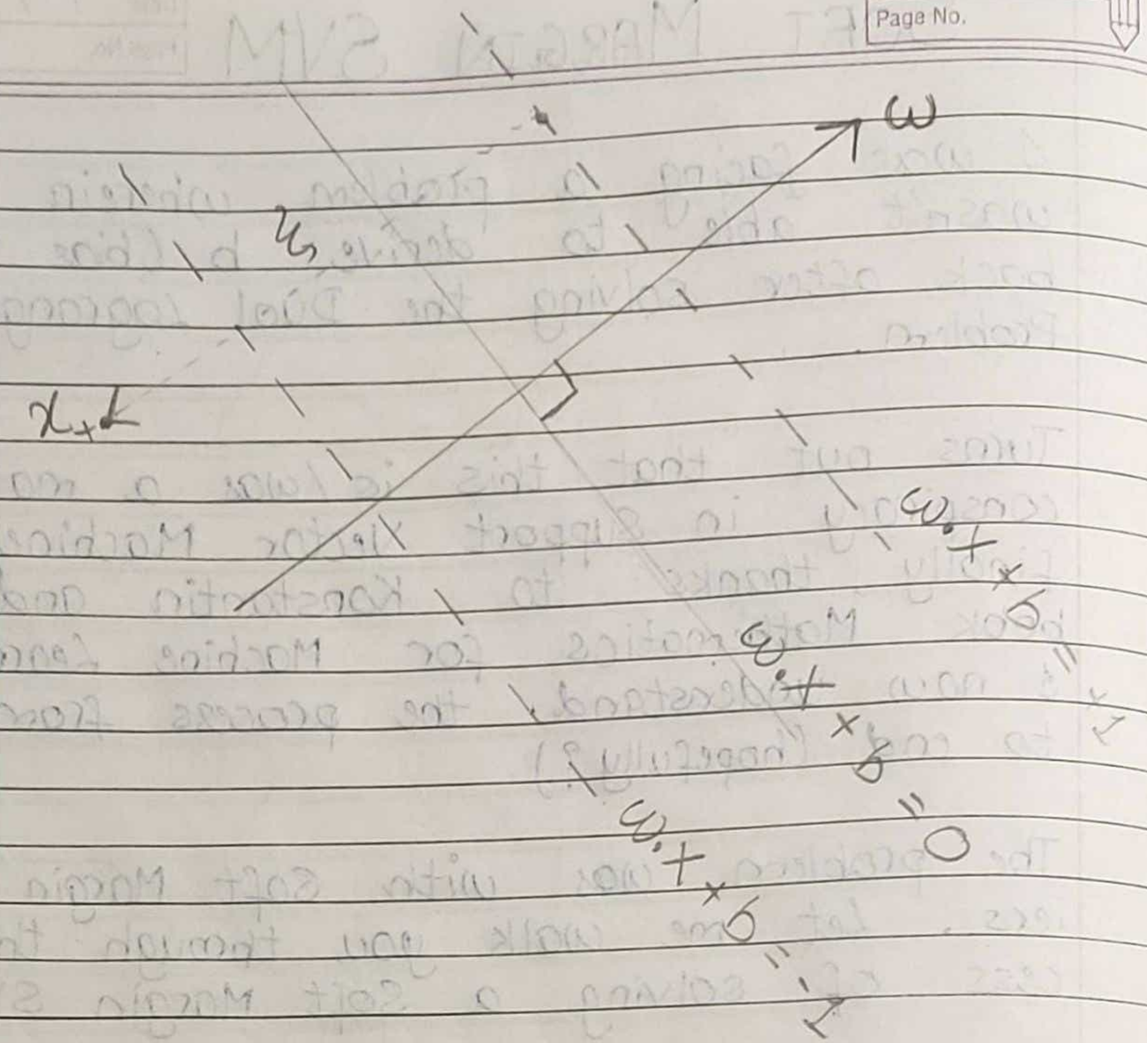
I was facing a problem wherein I wasn't able to derive b (bias term) back after solving the Dual Lagrangian Problem.

Turns out that this is/was a matter of conspiracy in Support Vector Machines and finally, thanks to Konstantin and the book Mathematics for Machine Learning, I now understand the process from start to end (hopefully?)

The problem was with Soft Margin Classifiers. Let me walk you through the process of solving a Soft Margin SVM.

The broad idea is to allow for some missclassifications in exchange for a wider and a better generalizing margin classifier.
In other words, our model will intentionally misclassify the noise/outliers in our imperfect data and only focus more on correctly classifying the data points which represent the true trend in data so that we can generalize well.

So we allow for some 'slack' in our model. Let's introduce a slack variable ξ_i for each training example. ξ_i will quantify the deviation a sample has from its respective margin.



So, here the slack variable ξ shows how far x_+ is from the + margin. Ideally, we want to minimize the slack and maximize the margin at the same time.

$$\therefore \min_{\omega, b, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{m} \xi_i$$

Since ξ_i is slack, it gets subtracted from the margin, constraining ξ_i to be non-negative (RHS is always +ve)

$$\therefore \quad y_i(\omega \cdot x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0 \qquad \forall \ i = 1, 2, \ldots, m$$

Date: / /
Page No.

Thus, our optimization problem becomes

$$\min_{\omega, b, \varepsilon} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} \varepsilon_i$$

subject to $y_i(\omega \cdot x_i + b) \geq 1 - \varepsilon_i$

$\varepsilon_i \geq 0 \quad \forall\ i = 1, 2, \ldots, m$

Forming the Primal Lagrangian Problem,

$$L_p(\omega, b, \varepsilon, \alpha, \gamma) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} \varepsilon_i$$

$$- \sum_{i=1}^{m} \alpha_i \left[y_i(\omega \cdot x_i + b) - 1 + \varepsilon_i \right] - \sum_{i=1}^{m} \gamma_i \varepsilon_i$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0 \quad \forall\ i = 1, 2, \ldots, m.$

Differentiating the Lagrangian w.r.t the 3 primal variables ω, b and ε:

$$\frac{\partial L}{\partial \omega} = \omega^T - \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^{m} \alpha_i y_i$$

$$\frac{\partial L}{\partial \varepsilon_i} = C - \alpha_i - \gamma_i$$

Setting each of these partial derivatives to zero, we get -

Date: / /
Page No.

$$\boxed{w = \sum_{i=1}^{m} \alpha_i y_i x_i} \quad - (1)$$ (Same as hard margin)

$$\boxed{\sum_{i=1}^{m} \alpha_i y_i = 0} \quad - (2)$$

$$\boxed{C = \alpha_i + \gamma_i} \quad - (3)$$

Substituting (1) into our primal Lagrangian

$$\Rightarrow \mathcal{L}_D(\varepsilon, \alpha, \gamma) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{i=1}^{m} y_i \alpha_i \left(\left\{ \sum_{j=1}^{m} y_j \alpha_j x_j \right\} \cdot x_i \right) + C \sum_{i=1}^{m} \varepsilon_i - b \underbrace{\sum_{i=1}^{m} y_i \alpha_i}_{0} + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i \varepsilon_i - \sum_{i=1}^{m} \gamma_i \varepsilon_i$$

$$\Rightarrow \mathcal{L}_D(\varepsilon, \alpha, \gamma) = -\frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{i=1}^{m} \alpha_i + \sum_{i=1}^{m} \underbrace{(C - \alpha_i - \gamma_i)}_{0} \varepsilon_i$$

$$\Rightarrow \mathcal{L}_D(\varepsilon, \alpha, \gamma) = -\frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{i=1}^{m} \alpha_i$$

We have eliminated ε from our dual problem. One less variable to worry about!
Even γ is no more in the Duality.

Date: / /
Page No.

Now, $\gamma_i \geq 0$ and $\alpha_i \geq 0$

$\Rightarrow$ Given eq ③,

$$C = \alpha_i + \gamma_i \quad \text{where } \gamma_i \geq 0$$

$\Rightarrow$ $\alpha_i \leq C$

$\therefore$ $\boxed{0 \leq \alpha_i \leq C}$

Dual SVM :

$$\min_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j)$$

$$\text{subject to } \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C \quad \forall \; i = 1, 2, \ldots, m$$

Now, once we solve for α, we end up with our Support Vectors for which $\alpha > 0$ ($\alpha \leq C$).

We can now recover the Primal parameters ω^* and b^* (controversial).

$$\omega^* = \sum_{i=1}^{m} \alpha_i y_i x_i \qquad \text{from } ⑦$$

The next step had got me, for b we can use the ω obtained in the equation:

$w^* \cdot x_i + b^* = y_i$ for some Support Vector

that lies "on" the margin's boundary so that $y_i = +1$ or -1.

$$\Rightarrow \boxed{b^* = y_i - w^* \cdot x_i}$$

BUT, there might be no examples (Support Vectors) that lie exactly on the magin. Theorectically, it's possible that all the SVs of a soft margin SVM are violating the margin.
[All SVs which have $0 < \alpha < C$ are necessarily on the margin]

Through some complicated math, we can say that in such cases we should calculate the median value of $|y_i - w^* \cdot x_i|$ $\forall\ i = 1, 2, \ldots, m$.

$$\rightarrow b^* = \text{median}\left(|y_i - w^*_i \cdot x_i|\right) \quad \forall\ i = 1, 2, \ldots, m.$$