ML Under Modern Optimization Lens - Sparse and Robust Classification - Exercícios

Giovanni Amorim

Junho, 2023

1 Formular e otimizar o problema primal do SVM com folga

Problema primal do SVM:

$$
\begin{aligned}
\min_{w,b,\epsilon} \quad & \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\epsilon_i \\
\text{s.a.} \quad & y_i(w^T x_i + b) \geq 1 - \epsilon_i \quad && \forall\, i \in \{1, ..., N\} \\
& \epsilon_i \geq 0 && \forall\, i \in \{1, ..., N\}
\end{aligned} \tag{1}
$$

A implementação foi feita adicionando a restrição de esparcidade utilizando técnica big-M:

$$
\begin{aligned}
\min_{w,b,\epsilon} \quad & \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\epsilon_i \\
\text{s.a.} \quad & y_i(w^T x_i + b) \geq 1 - \epsilon_i \quad && \forall\, i \in \{1, ..., N\} \\
& \epsilon_i \geq 0 && \forall\, i \in \{1, ..., N\} \\
& ||w||_0 \leq k
\end{aligned} \tag{2}
$$

Foram gerados dados aleatórios para avaliar a implementação: $X \in \mathbf{R}^{3000\times 20},\ y \in \{0,1\}^{3000}$. O vetor y foi gerado a partir de regras lógicas simples utilizando as features 1, 2, 5 e 6. As observações foram separadas em treino (70%) e teste (30%).

Na etapa de treino, foram utilizados os parâmetros $C = 1, k = 4$ e o vetor w encontrado foi:

1. $w_1 = -0.5077$
2. $w_2 = 1.2451$
3. $w_5 = 0.6610$
4. $w_6 = -0.6757$

Sendo $w_i = 0$ para todas as outras features. Logo, o modelo foi capaz de identificar as features relevantes. Foi avaliada a métrica de acurácia em treino (81.38%) e teste (80.66%).

2 Apresentar o dual do problema descrito

Relembrando o problema primal:

$$\begin{aligned} \min_{w,b,\epsilon} \quad & \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\epsilon_i \\ \text{s.a.} \quad & y_i(w^T x_i + b) \geq 1 - \epsilon_i \quad \forall\, i \in \{1, ..., N\} \\ & \epsilon_i \geq 0 \qquad\qquad\qquad\quad \forall\, i \in \{1, ..., N\} \end{aligned} \tag{3}$$

Vamos construir a função lagrangiana adicionando os multiplicadores α e β:

$$L(w, b, \epsilon, \alpha, \beta) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i[1 - \epsilon_i - y_i(w^T x_i + b)] - \sum_{i=1}^{N}\beta_i\epsilon_i$$

Seguiremos encontrando as derivdas da lagrangiana para os parâmetros w, b, ϵ:

$$\frac{\partial}{\partial w}L(w, b, \epsilon, \alpha, \beta) = w - \sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\frac{\partial}{\partial w}L(w, b, \epsilon, \alpha, \beta) = 0 \to w = \sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\frac{\partial}{\partial b}L(w, b, \epsilon, \alpha, \beta) = \sum_{i=1}^{N}\alpha_i y_i$$

$$\frac{\partial}{\partial b}L(w, b, \epsilon, \alpha, \beta) = 0 \to \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\frac{\partial}{\partial \epsilon_i}L(w, b, \epsilon, \alpha, \beta) = C - \alpha_i - \beta_i$$

$$\frac{\partial}{\partial \epsilon_i}L(w, b, \epsilon, \alpha, \beta) = 0 \to C - \alpha_i - \beta_i = 0$$

Como as variáveis duais α, β são não-negativas, temos que a igualdade equivale a:

$$0 \leq \alpha_i \leq C \; \forall\, i \in \{1, ..., N\}$$

Vamos fazer as substituições na função original:

$$L(w,b,\epsilon,\alpha,\beta) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i[1-\epsilon_i - y_i(w^T x_i + b)] - \sum_{i=1}^{N}\beta_i\epsilon_i$$

$$= \frac{1}{2}(\sum_{i=1}^{N}\alpha_i y_i x_i)^T(\sum_{i=1}^{N}\alpha_i y_i x_i) + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i[1-\epsilon_i - y_i((\sum_{j=1}^{N}\alpha_j y_j x_j)^T x_i + b)] - \sum_{i=1}^{N}\beta_i\epsilon_i$$

Expandindo:

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\epsilon_i - \sum_{i=1}^{N}\alpha_i y_i(\sum_{j=1}^{N}\alpha_j y_j x_j)^T x_i + b\sum_{i=1}^{N}\alpha_i y_i - \sum_{i=1}^{N}\beta_i\epsilon_i$$

Colocando os produtos no mesmo formato

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\epsilon_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + b\sum_{i=1}^{N}\alpha_i y_i - \sum_{i=1}^{N}\beta_i\epsilon_i$$

Somando os termos parecidos:

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + C\sum_{i=1}^{N}\epsilon_i + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\epsilon_i + b\sum_{i=1}^{N}\alpha_i y_i - \sum_{i=1}^{N}\beta_i\epsilon_i$$

Simplificando a partir das restrições $\sum_{i=1}^{N}\alpha_i y_i = 0$ e $C - \alpha_i - \beta_i = 0$:

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{N}\alpha_i$$

$$= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

Finalmente, escrevendo o problema dual:

$$\begin{aligned} \max_{\alpha\geq 0} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ \text{s.a.} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 & \forall\, i \in \{1, ..., N\} \\ & 0 \leq \alpha_i \leq C & \forall\, i \in \{1, ..., N\} \end{aligned} \tag{4}$$