2
Um pouco da teoria de otimização

Algumas definições importantes da teoria de otimização são introduzidas. Para maiores detalhes, veja Vapnik, 1998 (29) e Cristianini e Shawe-Taylor, 2000(7).

Definição 2.1 (Problema de otimização primal) *Considere as funções* $f; g_i, i = 1, ..., m;$ *e* $h_j, j = 1, ..., n,$ *definida em um domínio* $\Psi \subset \mathbb{R}^n$,

$$\begin{array}{lll} \textit{minimize} & f(x), & x \in \Psi \\ \textit{sujeito a} & g_i(x) \leq 0, & i = 1, ..., m \\ & h_j(x) = 0, & j = 1, ..., n \end{array}$$

onde $f(x)$ *é a função objetivo e as relações restantes são chamadas, respectivamente, restrições de desigualdade e de igualdade.*

A região do domínio onde a função objetivo está definida e onde todas as restrições são satisfeitas é denominada região viável e é denotada por:

$$R = \{x \in \Psi : g(x) \leq 0 \text{ e } h(x) = 0\}$$

Uma solução para o problema de otimização é um ponto $x^* \in \mathbb{R}$ tal que não existe outro ponto $x \in \mathbb{R}$ para o qual $f(x) < f(x^*)$. Tal ponto é chamado mínimo global. Um mínimo local é um ponto x^* para o qual $f(x^*) \leq f(x)$ para todo x com $\| x - x^* \| < \varepsilon$.

Um problema de otimização no qual a função objetivo, as restrições de igualdade e desigualdade são todas funções lineares é chamado um problema de programação linear. Caso a função objetivo seja quadrática e as restrições lineares, será denominado de problema de programação quadrática.

Definição 2.2 *Uma restrição de desigualdade* $g_i(x) \leq 0$ *é dita ser ativa se a solução ótima* x^**, satisfaz* $g(x^*) = 0$*, caso contrário é dita ser uma restrição inativa. As restrições de igualdade serão sempre ativas.*

Definição 2.3 (conjunto convexo) *Um conjunto* X *é chamado convexo se para todo* $x, x\prime \in X$ *e para todo* $\lambda \in [0, 1]$*, temos:* $\lambda x + (1 - \lambda)x\prime \in X$.

Definição 2.4 (função convexa) *Uma função f definida em um conjunto X, o qual não precisa ser convexo, é chamada convexa se e somente se para todo $x, x\prime \in X$ e para todo $\lambda \in [0,1]$ tal que $\lambda x + (1-\lambda)x\prime \in X$, temos $f(\lambda x + (1-\lambda)x\prime) \leq \lambda f(x) + (1-\lambda)f(x\prime)(*)$.*

Uma função f é chamada estritamente convexa se para $x \neq x\prime$ e $\lambda \in (0,1), (*)$ é a desigualdade estrita.

Se uma função f é convexa, qualquer mínimo local x^* de um problema de otimização irrestrito com função objetivo f é também um mínimo global, pois para qualquer $y \neq x^*$ por definição de mínimo local existe θ suficientemente próximo de 1 tal que

$f(x^*) \leq f(\theta x^* + (1-\theta)y) \leq \theta f(x^*) + (1-\theta)f(y)$. Logo, $f(x^*) \leq f(y)$.

Definição 2.5 *Um problema de otimização no qual o conjunto Ψ, a função objetivo e todas as restrições são convexas é dito ser convexo.*

Conforme será visto, treinar uma máquina SVR vai significar resolver um problema de otimização quadrática convexa.

2.1 Teoremas da teoria de otimização

Os principais resultados que serão utilizados para desenvolver soluções eficientes para a tarefa de otimizar SVRs são descritos. Maiores detalhes e demonstrações dos teoremas apresentados podem ser encontradas em (5).
O primeiro método analítico para resolver problemas de otimização foi desenvolvido por Fermat em 1629, o qual descreveu um método para obter máximos e mínimos de uma função sem restrições.

Teorema 2.6 (Fermat) *Considere uma função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1. Uma condição necessária para que $x^* = (x_1^*, ..., x_n^*)$ seja um mínimo para f é que $\frac{\partial f(x*)}{\partial x_i} = 0, i = 1, ...n$. Esta condição juntamente com a convexidade de f constitui uma condição suficiente.*

Em 1788, Lagrange sugeriu um método para resolver problemas de otimização com restrições de igualdade. Inicialmente, precisa-se definir uma função que incorpore tanto a função objetivo quanto às restrições, conhecida como Lagrangiana.

Definição 2.7 *Dado um problema de otimização com função objetivo $f(x)$ e restrições de igualdade $h_i(x) = 0, i = 1, ..., m$, definimos a função Lagrangiana como:*

$L(x, \beta) = f(x) + \sum_{i=1}^{m} \beta_i h_i(x)$ *onde os coeficientes β_i são chamados os multiplicadores de Lagrange.*

Teorema 2.8 (Lagrange) *Uma condição necessária para que um ponto $x^* = (x_1^*, ..., x_n^*)$ seja um mínimo para $f: \mathbb{R}^n \to \mathbb{R}$ sujeito a $h_i(x) = 0$, com f, h_i de classe C^1 é que as seguintes condições, chamadas condições de estacionariedade, sejam satisfeitas:*

$$\frac{\partial L(x^*, \beta^*)}{\partial x_i} = 0$$
$$\frac{\partial L(x^*, \beta^*)}{\partial \beta_i} = 0,$$

para algum valor de β^*. As condições anteriores serão também suficientes caso $L(x, \beta^*)$ seja uma função convexa de x.

Observação: Uma vez que as restrições são iguais a zero, o valor do Lagrangiano no ponto ótimo será igual ao valor da função objetivo, ou seja,

$$L(x^*, \beta^*) = f(x^*)$$

Considera-se o problema de otimização com restrições de igualdade e desigualdade e define-se o Lagrangiano generalizado.

Definição 2.9 *Dado um problema de otimização com domínio $\Psi \subset \mathbb{R}^n$,*

$$\begin{array}{lll} \textit{minimize} & f(x), & x \in \Psi \\ \textit{sujeito a} & g_i(x) \leq 0, & i = 1, ..., m \\ & h_j(x) = 0, & j = 1, ..., n \end{array}$$

Definimos a função de Lagrange generalizada como:

$$\begin{aligned} L(x, \alpha, \beta) &= f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{j=1}^{n} \beta_j h_j(x) \\ &= f(x) + \alpha' g(x) + \beta' h(x). \end{aligned}$$

Assim, o problema dual de Lagrange pode ser definido.

Definição 2.10 *O problema dual de Lagrange do problema primal da definição 2.1 é o seguinte problema:*

$$\begin{array}{ll} \textit{maximize} & \theta(\alpha, \beta), \\ \textit{sujeito a} & \alpha \geq 0, \end{array}$$

onde $\theta(\alpha, \beta) = \inf_{x \in \Psi} L(x, \alpha, \beta)$.

Teorema 2.11 (Dualidade Fraca) *Considere que $x \in \Psi$ seja uma solução viável do problema primal da definição 2.9 e (α, β) uma solução viável do problema dual da definição 2.10. Então $f(x) \geq \theta(\alpha, \beta)$.*

Corolário 2.12 *Seja x^*e (α^*, β^*) são soluções viáveis para os problemas primal e dual, respectivamente, e suponha que $f(x^*) = \theta(\alpha^*, \beta^*)$ onde $\alpha^* \geq 0$ e $g(x^*) \leq 0$, $h(x^*) = 0$, então x^* e (α^*, β^*) são soluções ótimas para o problema primal e dual, respectivamente.*

Teorema 2.13 (Dualidade Forte) *Se o problema primal P tem uma solução ótima x^* então o dual D tem uma solução ótima (α^*, β^*) tal que $f(x^*) = \theta(\alpha^*, \beta^*)$*

Em 1951, Kuhn e Tucker estenderam o método de multiplicadores de Lagrange para permitir restrições de desigualdades.

Teorema 2.14 (Kuhn-Tucker) *Considere o problema de otimização com domínio convexo $\Psi \subset R^n$,*

$$\begin{array}{lll} \textit{minimize} & f(x), & x \in \Psi \\ \textit{sujeito a} & g_i(x) \leq 0, & i = 1, ..., m \\ & h_j(x) = 0, & j = 1, ..., n \end{array}$$

Com f de classe C^1, convexa, g_i, h_j funções afins, as condições necessárias e suficientes para que x^* seja um ótimo é a existência de α^*, β^* tal que:

$$\begin{array}{ll} \frac{\partial L(x^*, \alpha^*, \beta^*)}{\partial x} = 0, & \\ \frac{\partial L(x^*, \alpha^*, \beta^*)}{\partial \beta} = 0, & \\ \alpha_i g_i(x) = 0, & i = 1, ..., m, \\ g_i(x) \leq 0, & i = 1, ..., m, \\ \alpha_i^* \geq 0, & i = 1, ..., m, \end{array}$$

A terceira condição é conhecida como condição complementar de Karush-Kuhn-Tucker (KKT).