MLCheatsheet

1. Bài toán Gốc: Toán học (Theory)

1. Thiết lập:
- $x, w \in \mathbb{R}^{M-1}$; Siêu phẳng: $w^T x + b = 0$.

2. Dẫn dắt: Max Margin $\Leftrightarrow$ Min $\|w\|$ $\Leftrightarrow$ Min $\frac{1}{2}\|w\|^2$.

3. Công thức Primal:

$$\boxed{\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 \\ &\text{s.t. } t_n(w^T x_n + b) \ge 1, \forall n \end{aligned}}$$

2. Bài toán Gốc: Implementation (CVXOPT)

Solver: $\min \frac{1}{2}u^T P u + q^T u$ s.t $Gu \le h$. **Biến:** $u = [\mathbf{w}; \mathbf{b}]_{(M\times 1)}$.

Mapping:

$$\boxed{\begin{aligned} &P = \text{diag}(\mathbf{1}, \dots, \mathbf{1}, \mathbf{0})_{(M\times M)} \\ &q = \mathbf{0}_{(M\times 1)} \\ &G = -[\text{diag}(\mathbf{t}) \cdot \mathbf{X}, \quad \mathbf{t}]_{(N\times M)} \\ &h = -\mathbf{1}_{(N\times 1)} \end{aligned}}$$

3. Lagrangian & Điều kiện KKT

1. Hàm Lagrangian:

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + \sum_{n=1}^{N} \underbrace{\alpha_n}_{\ge 0} \underbrace{\{1 - t_n(w^T x_n + b)\}}_{\substack{\text{đúng: } \le 0 \\ \text{sai: } > 0}}$$

$$= \frac{1}{2}\|w\|^2 + E_{\text{data}} \text{ (cost trên dữ liệu)}$$

Tính chất: Với $\alpha_n \ge 0$, (w^*, b^*) tối ưu:

$$\boxed{\mathcal{L}(w^*, b^*, \alpha) \le \frac{1}{2}\|w^*\|^2}$$

2. Tính đạo hàm:

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{n=1}^{N} \alpha_n t_n x_n$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{n=1}^{N} \alpha_n t_n$$

3. Điều kiện dừng (KKT-1):
- $\nabla_w \mathcal{L} = 0 \Rightarrow \boxed{w = \sum_{n=1}^{N} \alpha_n t_n x_n}$ (1)
- $\nabla_b \mathcal{L} = 0 \Rightarrow \boxed{\sum_{n=1}^{N} \alpha_n t_n = 0}$ (2)

4. Các điều kiện KKT khác:
- **(KKT-2)** Ràng buộc gốc: $1 - t_n(w^T x_n + b) \le 0$
- **(KKT-3)** Ràng buộc dual: $\alpha_n \ge 0$, $\forall n$
- **(KKT-4)** Điều kiện bù: $\alpha_n\{1 - t_n(w^T x_n + b)\} = 0$

Ý nghĩa: Thỏa KKT $\Rightarrow (w, b, \alpha) = (w^*, b^*, \alpha^*)$

5. Tiêu chuẩn Slater (cho SVM):

Nếu phần trong của tập khả thi không rỗng thì:

$$\exists (w, b) \text{ s.t. } 1 - t_i(w^T x_i + b) < 0, \ \forall i$$

$$\Longrightarrow \text{strong duality } (p^* = d^*) \ \Leftrightarrow \ \text{duality gap} = 0$$

4. Bài toán Dual: Biến đổi (Math)

Thay (1), (2) vào $\mathcal{L}$:

A. Thay w vào $\frac{1}{2}w^T w$:

$$\frac{1}{2}\left(\sum \alpha_i t_i x_i\right)^T \left(\sum \alpha_j t_j x_j\right) = \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j t_i t_j (x_i^T x_j) \ (*)$$

B. Thay w vào $-\sum \alpha_n t_n w^T x_n$:

$$-\sum_n \alpha_n t_n \underbrace{\left(\sum_m \alpha_m t_m x_m\right)^T}_{w^T} x_n = -2 \times (*)$$

C. Kết quả ($A + B = -A$):

$$\boxed{\begin{aligned} &\max_{\alpha} g(\alpha) = \sum_n \alpha_n - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j t_i t_j (x_i^T x_j) \\ &\text{s.t. } \alpha_n \ge 0; \quad \sum \alpha_n t_n = 0 (g(\alpha) \text{ luôn là hàm lồi}) \end{aligned}}$$

5. Bài toán Dual: Implementation (CVXOPT)

Solver: $\min \frac{1}{2}\alpha^T P \alpha + q^T \alpha$ s.t $G\alpha \le h, A\alpha = b$. **Biến:** $\alpha \in \mathbb{R}^N$.

Mapping:

$$\boxed{\begin{aligned} &P = \mathbf{K}_{\text{Gram}(N\times N)} \quad (K_{ij} = t_i t_j x_i^T x_j) \\ &q = -\mathbf{1}_{(N\times 1)} \quad (\text{Max } \Sigma \to \text{Min } -\Sigma) \\ &G = -I_{(N\times N)}; \ h = \mathbf{0}_{(N\times 1)} \\ &A = t^T; \ b = 0 \end{aligned}}$$

6. Soft Margin & Kernel

1. Soft Margin (ý tưởng): Cho phép một số điểm vi phạm margin bằng biến slack ξ_n, đổi lại bị phạt trong hàm mục tiêu.

2. Bài toán gốc (Primal – soft margin):

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \xi_n$$

$$\text{s.t. } \ t_n(w^T x_n + b) \ge 1 - \xi_n, \quad \xi_n \ge 0$$

Ý nghĩa C: C lớn $\Rightarrow$ phạt mạnh (ít sai, lề hẹp); C nhỏ $\Rightarrow$ cho sai nhiều (lề rộng).

3. Dạng đối ngẫu (Dual):

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\sum_n \xi_n - \sum_n \alpha_n\{t_n(w^T x_n + b) - 1 + \xi_n\} - \sum_n \mu_n \xi_n$$

$$\max_{\alpha} \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j t_i t_j \, x_i^T x_j$$

$$\text{s.t. } \boxed{0 \le \alpha_n \le C}, \quad \sum_{n=1}^{N} \alpha_n t_n = 0$$

Phân loại theo KKT:
- $\alpha_n = 0$: điểm ngoài margin (không support)
- $0 < \alpha_n < C$: support vector trên margin
- $\alpha_n = C$: support vector vi phạm / phân loại sai

4. Kernel Trick: Thay tích vô hướng:

$$x_i^T x_j \ \longrightarrow \ k(x_i, x_j)$$

$$f(x) = \sum_{n \in SV} \alpha_n t_n \, k(x_n, x) + b, \qquad y = \text{sign}(f(x))$$

Deep Learning

Goal: Multiclass classification ($y \in \{1, \dots, K\}$).

Model (Softmax):

$$p_{ik} = \frac{\exp(\mathbf{w}_k^\top \mathbf{x}_i + b_k)}{\sum_{j=1}^{K} \exp(\mathbf{w}_j^\top \mathbf{x}_i + b_j)}$$

Loss (Categorical Cross-Entropy):

$$\mathcal{L}_{\text{CE}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} y_{ik} \log p_{ik} = -\frac{1}{n}\sum_{i=1}^{n} \log p_{i,y_i}$$

Prediction:

$$\hat{y}_i = \arg\max_k p_{ik}$$

Notes:
- $y_{ik} \in \{0, 1\}$: nhãn one-hot của mẫu i tại lớp k
- p_{i,y_i}: xác suất dự đoán của lớp đúng của mẫu i

Vectorized form:
- **Vector (single sample):** for $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{M\times N}$, $\mathbf{b} \in \mathbb{R}^M$
  $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b} \in \mathbb{R}^M$
- **Matrix (mini-batch):** for $\mathbf{X} \in \mathbb{R}^{B\times N}$ (rows are samples)
  $\mathbf{Y} = \mathbf{X}\mathbf{W}^\top + \mathbf{1}\mathbf{b}^\top \in \mathbb{R}^{B\times M}$ (broadcast $\mathbf{b}$ to all rows)

Activation functions:

Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \qquad \sigma'(x) = \sigma(x)(1 - \sigma(x))$$

Tanh

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \qquad \tanh'(x) = 1 - \tanh^2(x)$$

ReLU

$$\text{ReLU}(x) = \max(0, x), \qquad \text{ReLU}'(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x \le 0 \end{cases}$$

Leaky ReLU

$$\text{LReLU}(x) = \begin{cases} x, & x \ge 0 \\ \alpha x, & x \le 0 \end{cases}, \qquad \text{LReLU}'(x) = \begin{cases} 1, & x \ge 0 \\ \alpha, & x \le 0 \end{cases}$$

SiLU (Swish)

$$\text{SiLU}(x) = x\,\sigma(x), \qquad \text{SiLU}'(x) = \sigma(x) + x\,\sigma(x)(1 - \sigma(x))$$

Linear Regression Solution:

Model:

$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b$$

Loss (MSE):

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2$$

Closed-form Solution (Normal Equation):

$$\mathbf{w}^* = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

With regularization (Ridge):

$$\mathbf{w}^* = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^\top \mathbf{y}$$