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:

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

2. Bài toán Gốc: Implementation (CVXOPT)

Solver: min $\frac{1}{2}u^T P u + q^T u$ s.t $Gu \leq h$. **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:

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

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

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

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

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: $1 - t_n(w^T x_n + b) \leq 0$
- **(KKT-3)** Ràng buộc dual: $\alpha_n \geq 0$, $\forall n$
- **(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 ($A + B = -A$):

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

5. Bài toán Dual: Implementation (CVXOPT)

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

Mapping:

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

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):

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

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

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):

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

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

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

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:

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

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

1. Giới thiệu & Mục tiêu (Introduction)

1. Đầu vào:
- Ma trận dữ liệu X có kích thước $N \times D$ (N mẫu, D chiều). Thường D rất lớn.
- Độ phức tạp tính toán tăng lên: ma trận Kgram, kernel, hiệp phương sai, khoảng cách giữa các điểm dữ liệu, số nơron lớp ẩn tăng. Và cần mô hình phức tạp hơn, điểm dữ liệu lớn hơn dễ gây overfit.

2. Mục tiêu:
- Giảm số chiều $D \to M$ sao cho $M \ll D$.
- Đặc trưng mới không tương quan (uncorrelated).
- Giữ lại thông tin quan trọng nhất (Variance lớn nhất).

2. Kiến thức Toán nền tảng (Math Basis)

1. Hình chiếu: Chiếu vector x lên vector u: $l = \frac{u^T x}{\|u\|}$.

2. Dạng bậc hai & Đạo hàm:
- $u = [u_1, u_2, \dots, u_N]^T$, $v = [v_1, v_2, \dots, v_M]^T$.
- $A = [a_{ij}]_{N \times M}$.
- $u^T A v = \sum_{i=1}^{N} \sum_{j=1}^{M} a_{ij} u_i^T v_j$. ($aij$ và $u_i^T v_j$ là vô hướng).
- $\frac{\partial u^T A u}{\partial u} = 2Au$ (với A đối xứng).

3. Eigenvectors & Eigenvalues:

$$Au = \lambda u \quad \Leftrightarrow \quad (A - \lambda I)u = 0$$

(3.7)

$$=> \det(A - \lambda I) = 0$$

(3.8) Giải 3.8 để tìm λ_i (có tối đa N eigenvalues), thay vào 3.7 để tìm u_i.

4. Tính chất eigenvectors và eigenvalues S: Vuông, đối xứng, bán định dương ($u^T S u \geq 0$).

3. Cơ sở lý luận (Max Variance Theory)

1. Bài toán tối ưu: Tìm hướng u để phương sai khi chiếu dữ liệu lên đó là lớn nhất.

$$\boxed{u_1 = \underset{u}{\operatorname{argmax}}\, u^T S u \quad \text{s.t. } u^T u = 1}$$

2. Giải bằng Lagrangian: $\mathcal{L}(u, \lambda) = u^T S u - \lambda(u^T u - 1)$.

$$\frac{\partial \mathcal{L}}{\partial u} = 0 \Rightarrow 2Su - 2\lambda u = 0 \Rightarrow Su = \lambda u$$

3. Kết luận: Hướng u tối ưu chính là **eigenvector** ứng với **eigenvalue** lớn nhất của ma trận S.

4. Giải thuật PCA (Algorithm)

1. Chuẩn hóa (Centering): Tính vector trung bình $\mu = \frac{1}{N} \sum x_n$.

$$Z = X - \mu^T$$

2. Tính Covariance Matrix:

$$S = \frac{1}{N} Z^T Z$$

3. Phân rã (Eigendecomposition): Tìm u_k, λ_k của S. Sắp xếp $\lambda_1 \geq \lambda_2 \dots$. Chọn M vectors đầu $\to$ Ma trận U_M.

4. Phép chiếu (Rotation):

$$\boxed{X_{pca} = Z U_M}$$

5. Phục hồi (Reconstruction): $\hat{X} = X_{pca} U_M^T + \mu^T$ (Xấp xỉ).

5. Mối quan hệ với SVD

1. Định nghĩa SVD: Phân rã ma trận gốc X (đã center):

$$X = U \Sigma V^T$$

2. Liên hệ PCA:
- Eigenvectors của $S \approx X^T X$ chính là các cột của V.
- Eigenvalues $\lambda_i = \frac{s_i^2}{N}$ (s_i là singular values).

3. Ưu điểm: SVD ổn định hơn về mặt số học (numerical stability) so với việc tính trực tiếp S.