Foundations of Machine Learning
AI2000 and AI5000

FoML-35
Support Vector Machines (cntd.)
Duality to obtain the max margin classification

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence
& Learning Lab

So far in FoML

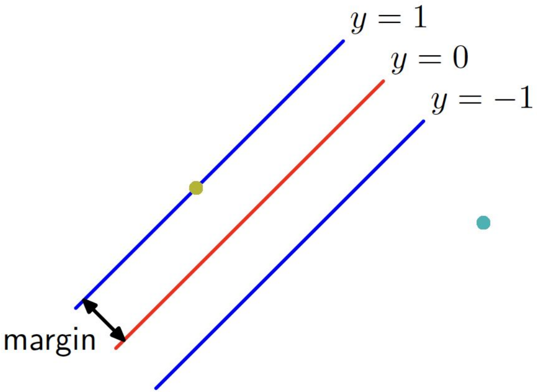
- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions
    b. Bias-Variance Decomposition
    c. Decision Theory - three broad classification strategies
    d. Neural Networks
- Unsupervised learning
    a. K-Means, Hierarchical, and GMM for clustering
- Kernelizing linear Models
    a. Dual representation, Kernel trick

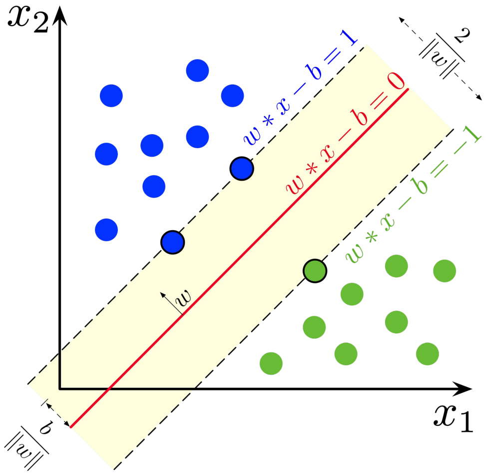
For today

- SVM (cntd.)
    - Duality to obtain the max margin classification

Max margin classifier

$$\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \qquad t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right) \geqslant 1, \qquad n=1,\ldots,N.$$





Max margin classifier

$$\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \qquad t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right) \geqslant 1, \qquad n=1,\ldots,N.$$

- Primal Lagrangian

$$L(\mathbf{w},b,\mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b)-1\right\}$$

Max margin classifier

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{ t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1 \right\}$$

KKT conditions

$$\begin{aligned}
&t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \geq 0 && \text{for} \quad n = 1, \ldots, N \\
&a_n \geq 0 && \text{for} \quad n = 1, \ldots, N \\
&a_n(t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1) = 0 && \text{for} \quad n = 1, \ldots, N
\end{aligned}$$

Max margin classifier

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{ t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1 \right\}$$

KKT conditions

$$\begin{aligned} &t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \geq 0 && \text{for} \quad n = 1, \ldots, N \\ &a_n \geq 0 && \text{for} \quad n = 1, \ldots, N \\ &a_n(t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1) = 0 && \text{for} \quad n = 1, \ldots, N \end{aligned}$$

Derive the dual Lagrangian via $\frac{\partial L}{\partial \mathbf{w}} = 0, \quad \frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \tilde{L}(\mathbf{a}) = \min_{\mathbf{x}, b} L(\mathbf{x}, b, \mathbf{a})$

Max margin classifier

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1\right\}$$

KKT conditions

$$\begin{aligned} & t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \geq 0 && \text{for} \quad n = 1,\ldots,N \\ & a_n \geq 0 && \text{for} \quad n = 1,\ldots,N \\ & a_n(t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1) = 0 && \text{for} \quad n = 1,\ldots,N \end{aligned}$$

Derive the dual Lagrangian via $\frac{\partial L}{\partial \mathbf{w}} = 0, \quad \frac{\partial L}{\partial b} = 0 \Longrightarrow \tilde{L}(\mathbf{a}) = \min_{\mathbf{x},b} L(\mathbf{x}, b, \mathbf{a})$

Now, solve for a* $\mathbf{a}^* = \arg\max_{\mathbf{a}} \tilde{L}(\mathbf{a})$

Max margin classifier

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{ t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1 \right\}$$

KKT conditions

$$\begin{array}{lll} t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \geq 0 & \text{for} & n = 1, \ldots, N \\ a_n \geq 0 & \text{for} & n = 1, \ldots, N \\ a_n(t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1) = 0 & \text{for} & n = 1, \ldots, N \end{array}$$

Derive the dual Lagrangian via $\frac{\partial L}{\partial \mathbf{w}} = 0, \quad \frac{\partial L}{\partial b} = 0 \Rightarrow \tilde{L}(\mathbf{a}) = \min_{\mathbf{x}, b} L(\mathbf{x}, b, \mathbf{a})$

Now, solve for a* $\mathbf{a}^* = \arg\max_{\mathbf{a}} \tilde{L}(\mathbf{a})$ then, solve for w*, b* $\mathbf{w}^*, b^* = \arg\min_{\mathbf{w}, b} L(\mathbf{w}, b, \mathbf{a}^*)$

Max margin classifier

- Let's form the dual Lagrangian for $L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{ t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1 \right\}$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w}^T - \sum_{n=1}^{N} a_n t_n \mathbf{x}_n^T = 0 \quad \Longrightarrow \quad \mathbf{w} = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n$$

$$\frac{\partial L}{\partial b} = -\sum_{n=1}^{N} a_n t_n = 0 \quad \Longrightarrow \quad \sum_{n=1}^{N} a_n t_n = 0$$

Eliminate w and b from L

Max margin classifier

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1\right\}$$

Applying the stationarity conditions

$$\mathbf{w} = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n \qquad \sum_{n=1}^{N} a_n t_n = 0$$

$$\tilde{L}(\mathbf{a}) =$$

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m \mathbf{x}_n^T \mathbf{x}_m$$

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m)$$

Max margin classifier

- Dual representation of the max margin (maximize w.r.t **a**)

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m \mathbf{x}_n^T \mathbf{x}_m$$

$$a_n \geq 0 \;\; \forall n = 1, \ldots N \qquad \sum_{n=1}^{N} a_n t_n = 0$$

Max margin classifier

- New prediction $y(\mathbf{x}_n) = \mathbf{w}^T\mathbf{x}_n + b \implies y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n^T \mathbf{x} + b$

Max margin classifier

- New prediction $y(\mathbf{x}_n) = \mathbf{w}^T \mathbf{x}_n + b$ $\Longrightarrow$ $y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n^T \mathbf{x} + b$

$$
\begin{aligned}
& t_n(\mathbf{w}^T \mathbf{x}_n + b) - 1 \geq 0 && \text{for} \quad n = 1, \ldots, N \\
& a_n \geq 0 && \text{for} \quad n = 1, \ldots, N \\
& a_n(t_n(\mathbf{w}^T \mathbf{x}_n + b) - 1) = 0 && \text{for} \quad n = 1, \ldots, N
\end{aligned}
$$

Max margin classifier

- New prediction $y(\mathbf{x}_n) = \mathbf{w}^T\mathbf{x}_n + b$ $\Longrightarrow$ $y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n^T \mathbf{x} + b$

$$
\begin{aligned}
&t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \geq 0 && \text{for} \quad n = 1,\ldots,N \\
&a_n \geq 0 && \text{for} \quad n = 1,\ldots,N \\
&a_n(t_n(\mathbf{w}^T\mathbf{x}_n + b) - 1) = 0 && \text{for} \quad n = 1,\ldots,N
\end{aligned}
$$

- Consider a_n
  - $>0 \rightarrow$
  - $=0 \leftarrow$

Max margin classifier

- New prediction $y(\mathbf{x}_n) = \mathbf{w}^T\mathbf{x}_n + b$ $\Rightarrow$ $y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n^T \mathbf{x} + b$

$\Rightarrow$ $y(\mathbf{x}) = \sum_{m \in S} a_m t_m k(\mathbf{x}_m, \mathbf{x}) + b$

- Find b using $t_n y_n(\mathbf{x}) = 1$ for support vectors

$$t_n \left(\sum_{m \in S} a_m t_m k(\mathbf{x}_m, \mathbf{x}_n) + b \right) = 1$$

$$\sum_{m \in S} a_m t_m k(\mathbf{x}_m, \mathbf{x}_n) + b = t_n$$

Next

- Gaussian Processes