Foundations of Machine Learning
AI2000 and AI5000

FoML-34
Support Vector Machines (cntd.)
Optimization with inequality constraints

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.)
    - Optimization with inequality constraints

SVM for binary classification

$$\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

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

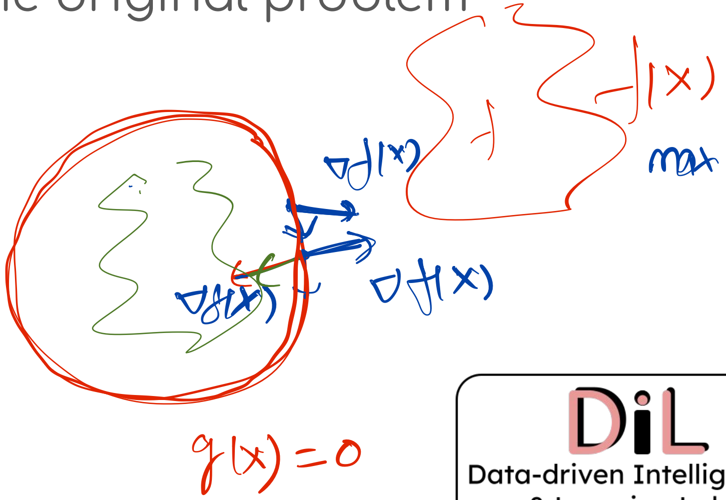
Constrained optimization (Quadratic programming) problem

Optimization with inequality constraints

Earlier - equality constraints

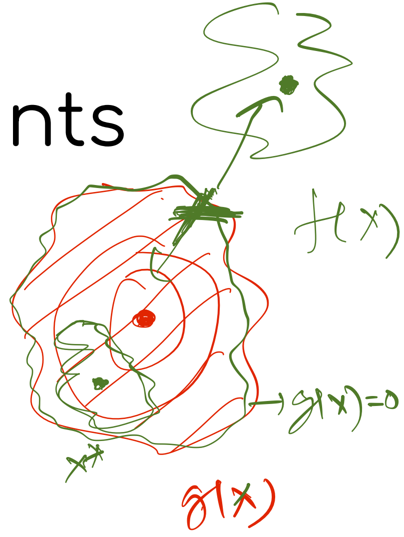
- Maximize f(x) with constraints g(x)=0
- We exploited: gradients are normal to the levelset g(x)=0
- → introduced a Lagrangian function L (x, λ)
- Stationary points of L → solution to the original problem

$$\nabla f(x) = \lambda \nabla g(x)$$



Optimization with inequality constraints

- Maximize f(x) such that g(x) ≥ 0
- Two possibilities
  a. Stationary point lies in region g(x) ≥ 0 (inactive constraints)
     - → $\nabla f(x) = -\lambda \nabla g(x)$ $\lambda > 0$ ✓
  b. Stationary point lies on the boundary g(x) = 0 (active constraints)
     - → $\lambda = 0$ ✓



Primal Lagrangian $$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

$\lambda \geq 0$

Optimization with inequality constraints

- Maximize f(x) such that g(x) ≥ 0
- Can be formulated as a max-min optimization problem

$$\max_{\mathbf{x}} \min_{\lambda} L(\mathbf{x}, \lambda) \text{ s.t. } \lambda \geq 0,\ g(\mathbf{x}) \geq 0,\ \lambda g(\mathbf{x}) = 0$$

Optimization with inequality constraints

- Maximize f(x) such that g(x) ≥ 0
- Can be formulated as a max-min optimization problem

$$\max_{\mathbf{x}} \min_{\lambda} L(\mathbf{x}, \lambda) \text{ s.t. } \lambda \geq 0,\ g(\mathbf{x}) \geq 0,\ \lambda g(\mathbf{x}) = 0$$

- Idea is to solve a dual Lagrangian (optimize w.r.t primal variable x for fixed values of λ)

$$\tilde{L}(\lambda) = \max_{\mathbf{x}} L(\mathbf{x}, \lambda) \text{ with } L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

$$f(x^*) \leq L(x^*, \lambda) \leq \tilde{L}(\lambda^*)$$

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

DiL
Data-driven Intelligence & Learning Lab

Optimization with inequality constraints

$$\tilde{L}(\lambda) = \max_{\mathbf{x}} L(\mathbf{x}, \lambda) \text{ with } L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

- Work the dual Lagrangian analytically
    - Stationarity condition ($\nabla \cancel{f} L(x) = 0$) eliminates x
    - → function of λ $\tilde{L}(\lambda)$
    - This forms an upper bound on the ~~primal max-min~~ problem (as a function of λ)
    - Minimize w.r.t. λ

$f(x^*)$

Optimization with inequality constraints

- Duality gap
  - For x' that satisfies g(x') ≥ 0, we have $f(\mathbf{x}') \leq L(\mathbf{x}', \lambda) \leq \tilde{L}(\lambda)$

$$\Rightarrow \quad \mathbf{p}^* = \max_{\mathbf{x}, g(\mathbf{x}) \geq 0} f(\mathbf{x}) \leq \min_{\lambda} \tilde{L}(\lambda) = \mathbf{d}^*$$

$(p^* - d^*)$

Optimization with inequality constraints

- Duality gap
  - For x' that satisfies g(x') ≥ 0, we have $f(\mathbf{x}') \leq L(\mathbf{x}', \lambda) \leq \tilde{L}(\lambda)$

$$\Rightarrow \quad \mathbf{p}^* = \max_{\mathbf{x}, g(\mathbf{x}) \geq 0} f(\mathbf{x}) \leq \min_{\lambda} \tilde{L}(\lambda) = \mathbf{d}^*$$

Most convex problems exhibit strong duality, i.e., p* = d*

Summary

- Primal problem maximize f(x) subject to g(x)≥0

$$\Rightarrow \quad \max_{\mathbf{x}} \min_{\lambda} L(\mathbf{x}, \lambda) \text{ s.t. } \lambda \geq 0,\ g(\mathbf{x}) \geq 0,\ \lambda g(\mathbf{x}) = 0$$

- Dual problem (find the lowest upper bound) $\min_{\lambda} \tilde{L}(\lambda)$ **subject to** $\lambda \geq 0$

Summary

$f(x^*) \leq \tilde{L}(\lambda)$

- Primal problem maximize f(x) subject to g(x)≥0

$$\Rightarrow \quad \max_{\mathbf{x}} \min_{\lambda} L(\mathbf{x}, \lambda) \text{ s.t. } \lambda \geq 0,\ g(\mathbf{x}) \geq 0,\ \lambda g(\mathbf{x}) = 0$$

primal problem

- Dual problem (find the lowest upper bound) $\min_{\lambda} \tilde{L}(\lambda)$ **subject to** $\lambda \geq 0$
- Steps
  - Define Lagrangian $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$
  - Compute the dual $\tilde{L}(\lambda)$
  - Solve the dual problem $\lambda^* = \min_{\lambda} \tilde{L}(\lambda)$ **subject to** $\lambda \geq 0$
  - Maximize the primal Lagrangian $\mathbf{x}^* = \arg\max_{\mathbf{x}} L(\mathbf{x}, \lambda^*)$

Next

- Kernel SVM