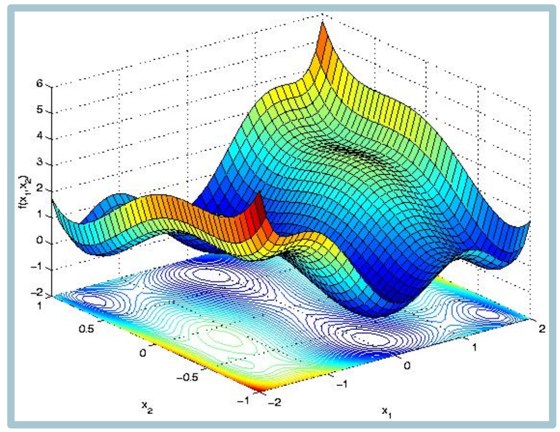


An Introduction to Optimization with Application to Machine Learning

Hamed Valizadegan

University of Pittsburgh

Motivation: Machine Learning

- Linear Regression

$$\underset{w,b}{\text{minimize}} \quad \sum_{i=1}^{n} \|w^t x_i + b - y_i\|^2$$

- SVM

$$\underset{w,b}{\text{minimize}} \quad \|w\|^2 + C \sum_{i=1}^{n} \varepsilon_i$$
$$\textit{subject to} \quad y_i(w^T x_i + b) \geq 1 - \varepsilon_i \;\; i = 1, \dots n$$
$$\varepsilon_i \geq 0 \;\; i = 1, \dots, n$$

- PGDM metric learning

$$\underset{P}{\text{minimize}} \quad \sum_{(x_i,\, x_j) \in S} \left\|x_i - x_j\right\|_P$$
$$\textit{subject to} \; \sum_{(x_i,\, x_j) \in D} \left\|x_i - x_j\right\|_P \geq 1$$
$$P \succcurlyeq 0$$

Optimization Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \textit{subject to} \quad & f_i(x) \leq 0 \;\; i = 1, \dots m \\ & h_i(x) = 0 \;\; i = 1, \dots, p \end{aligned}$$

- $x \in R^n$ is the variable to find
- $f_0: R^n \rightarrow R$ is called the objective (cost or utility) function
- $f_i: R^n \rightarrow R, i = 1, \dots m$ are the inequality constraints (defines a set)
- $h_i: R^n \rightarrow R, i = 1, \dots p$ are the equality constraints (defines a set)

- Solution: $p^* = \inf\{f_0(x) | f_i(x) \leq 0 \;\; i = 1, \dots m, h_i(x) = 0 \;\; i = 1, \dots, p\}$
- Constrained vs. unconstrained problems: whether you have the constrains or not.

- A feasible point x is optimal if $f_0(x) = p^*$; X_{OPT} is the set of optimal points.

Feasibility

- An optimization problem is feasible
  - if $x \in dom\ f_0$ (implicit constraints) and it satisfies all the (explicit) constraints $f_i(x) \leq 0\ \ i = 1, \dots m$ & $h_i(x) = 0\ \ i = 1, \dots, p$.
- For infeasible problems, we say $p^* = +\infty$

- Feasibility problem

$$\begin{aligned} &find && x \\ &subject\ to && f_i(x) \leq 0\ \ i = 1, \dots m \\ & && h_i(x) = 0\ \ i = 1, \dots, p \end{aligned}$$

- Equivalent to the following optimization problem

$$\begin{aligned} &minimize\ \ 0 && \\ &subject\ to && f_i(x) \leq 0\ \ i = 1, \dots m \\ & && h_i(x) = 0\ \ i = 1, \dots, p \end{aligned}$$

Locally Optimal Points

- For the following problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ s.t. \quad & f_i(x) \leq 0 \;\; i = 1, \ldots m \\ & h_i(x) = 0 \;\; i = 1, \ldots, p \end{aligned}$$

- x is locally optimum if there is an R > 0 such that x is optimal for the following problem

$$\begin{aligned} \underset{z}{\text{minimize}} \quad & f_0(z) \\ s.t. \quad & f_i(z) \leq 0 \;\; i = 1, \ldots m \\ & h_i(z) = 0 \;\; i = 1, \ldots, p \\ & \|z - x\|_2 \leq R \end{aligned}$$

Regularization

- A form of limiting the feasible search space of an optimization problem
- Can be considered as the prior information that the solution is located in the neighborhood of point x

$$
\begin{aligned}
&\underset{x}{\text{minimize}} \quad f_0(x) \\
&s.t. \;\; f_i(x) \le 0 \;\; i = 1, \dots m \\
&\qquad\qquad h_i(x) = 0 \;\; i = 1, \dots, p
\end{aligned}
\quad \rightarrow \quad
\begin{aligned}
&\underset{z}{\text{minimize}} \quad f_0(z) \\
&s.t. \;\; f_i(z) \le 0 \;\; i = 1, \dots m \\
&\qquad h_i(z) = 0 \;\; i = 1, \dots, p \\
&\qquad \|z - x\|_p \le R
\end{aligned}
$$

- Leads to sparse solution for x =0 and small p
- I will get back to this.

Convexity

- An optimization problem is convex if
  - $f_0: R^n \to R$ is a convex function
  - Constrains $f_i(x) \leq 0 \;\; i = 1, \dots m$ & $h_i(x) = 0 \;\; i = 1, \dots, p$ are convex sets.
  - $f_0: R^n \to R$, $f_i: R^n \to R, i = 1, \dots m$, $h_i: R^n \to R, i = 1, \dots p$ can be linear or nonlinear

- Importance
  - Any local optimum is a global optimum
  - Local optimality can be verified. No general tractable global optimum test
  - So, for convex problems, it is easy to check if a point is a global optimum.

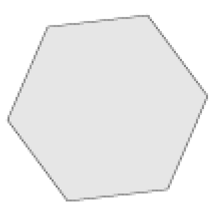
- Feasible set of a convex optimization problem is convex.
- Convex set and convex function??

Affine and Convex Sets

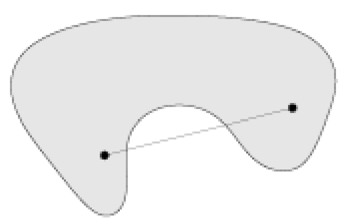
- Affine sets: the line through any two disjoint points
  - $x = \theta x_1 + (1-\theta)x_2, \quad \theta \in \mathbb{R}$
  - Or equivalently, solution set of linear equation $\{x | Ax = b\}$

- Line segment: line segment between two points
  - $x = \theta x_1 + (1-\theta)x_2, \quad 0 \le \theta \le 1$

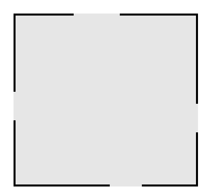
- Convex Sets: a set that contains the line segment of any two points of the set
  - $x_1, x_2 \in S, 0 \le \theta \le 1 \;\Longrightarrow\; \theta x_1 + (1-\theta)x_2 \in S$



Convex



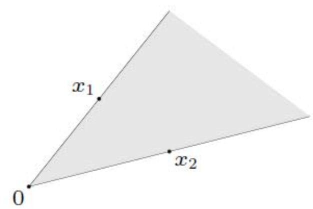
Non-convex

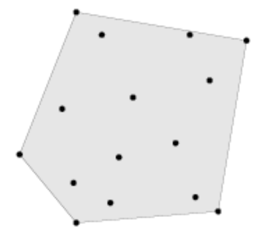


Non-Convex

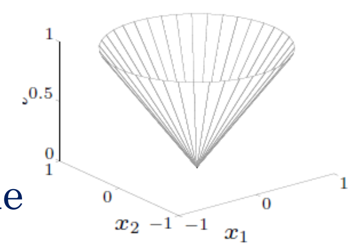
Convex Sets (examples)

- Convex hull of set $S = \{x_1, x_2, \ldots, x_k\}$: Set of all convex combinations of points in S
  - $\left\{x \middle| \sum_{i=1}^{k} \theta_i x_i \,,\, \sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0\right\}$





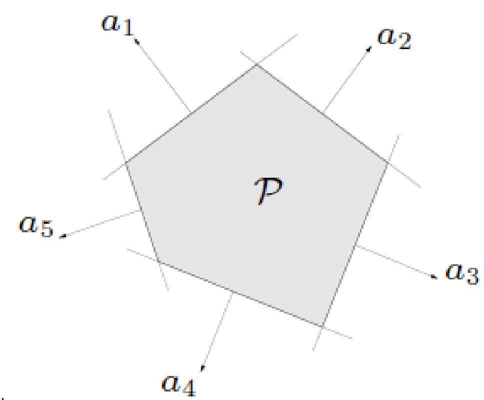
- Conic combination of two points
  - $x = \theta_1 x_1 + \theta_2 x_2, \quad 0 \leq \theta_1, \theta_2$
- Convex cone of set S: a set that contains all conic combinations of points in S
- Hyperplanes ($a^T x + b = 0$, linear equality)
- Halfspaces ($a^T x + b \leq 0$, linear inequality)
- Euclidean balls and Ellipsoids: $\{x |\ (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$ (P $\in \boldsymbol{S}^n{}_{++}$, i.e. P is positive-definite P)
- Norm ball: $\left\{x \middle| \|x - x_c\| \leq r\right\}$
- Norm cone: C=$\left\{(x, t) \middle| \|x\| \leq t\right\} \in \mathbb{R}^{n+1}$
  - Euclidean norm cone ($\|x\|_2$) is called second order cone



Operations that preserve convexity

- Intersection of convex sets
- The image of a convex set under affine (linear) function
  - $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$: F(x)=Ax+b
  - scaling (aS), translation(S+a), projection
- Perspective function
  - $F: \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$: F(x,t)=x/t,   dom(F)={(x,t) | t>0)
  - Image and inverse image of convex sets under perspective are convex
- Linear-fractional functions:
  - $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$: F(x,t)=$\frac{Ax+b}{c^Tx+d}$,   dom(F)={x | $c^Tx + d$>0)
  - Image and inverse image of convex sets under linear-fractional functions are convex

Convexity preserving operations (cont.)



- Intersection of convex sets is convex.
- Polyhedra is convex
  - Intersection of finite number of halfspaces and hyperplanes
- Positive semidefinite (PSD) cone: Set of all PSD matrices is convex
  - Intersection of infinite number of halfspaces and hyperspaces passing through origin ($\bigcap_{z \neq 0} \{X \in \boldsymbol{S}^n \mid z^T X z \geq 0\}$)
  - We denote it by $\boldsymbol{S}^n{}_+$

Generalized Inequalities

- Definition: A cone $K \subseteq \mathbb{R}^n$ is a proper cone if
  - K is convex
  - K is closed
  - K is solid: it has nonempty interior
  - K is pointed: it contains no line

- Generalized inequalities: defined by a proper cone K, is a partial ordering

$$x \preccurlyeq_K y \quad \Leftrightarrow y - x \in K$$

$$x \prec_K y \quad \Leftrightarrow y - x \in int\ K\ (interior\ of\ K)$$

- Examples
  - Componentwise inequality:

$$x \prec_{R_+{}^n} y \quad \Leftrightarrow y_i \geq x_i$$

  - Matrix inequality

$$X \prec_{\boldsymbol{s}_+{}^n} Y \quad \Leftrightarrow Y - X\ is\ PSD$$

Dual Cones

- Dual cone of a cone K: $K^* = \{y \mid y^T x \geq 0 \; for \; all \; x \in K\}$

  $$x \preccurlyeq_K y \iff y - x \in K$$

  $$x \prec_K y \iff y - x \in int\, K \; (interior \; of \; K)$$

- Examples
  - $K = {R_+}^n$: $K^* = {R_+}^n$
  - $K = {\boldsymbol{S}_+}^n$: $K^* = {\boldsymbol{S}_+}^n$, $(tr(XY) \geq 0)$
  - $\mathrm{K} = \{(x,t) \mid \|x\|_2 \leq t\}$: K^*=$\{(x,t) \mid \|x\|_2 \leq t\}$
  - $\mathrm{K} = \{(x,t) \mid \|x\|_1 \leq t\}$: K^*=$\{(x,t) \mid \|x\|_\infty \leq t\}$

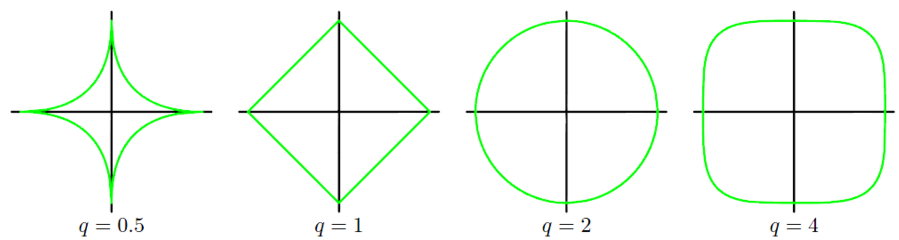
Convex Functions

- Definition: function $f(x)$: $\mathbb{R}^n \rightarrow \mathbb{R}$ is convex if the graph of the function lies between the line segment joining any two points of the graph.

- Formally: $f(x)$: $\mathbb{R}^n \rightarrow \mathbb{R}$ is convex if dom(f) is convex and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

- Examples in $\mathbb{R}$:
  - affine, exponential, powers ($x^\alpha, \alpha \leq 0 \; or \; \alpha \geq 1$), power of absolute value ($|x|^\alpha, \alpha \geq 1$)
- Example on $\mathbb{R}^n$
  - Norm$\|x\|_\alpha = (\sum_{i=1}^{n} |x_i|^\alpha)^{1/\alpha}, \alpha \geq 1$



- Example on $\mathbb{R}^{n \times m}$
  - Affine function $\mathrm{tr}(A^T X) + b = \sum_{i=1}^{m} \sum_{j-1}^{n} A_{ij} X_{ij} + b$

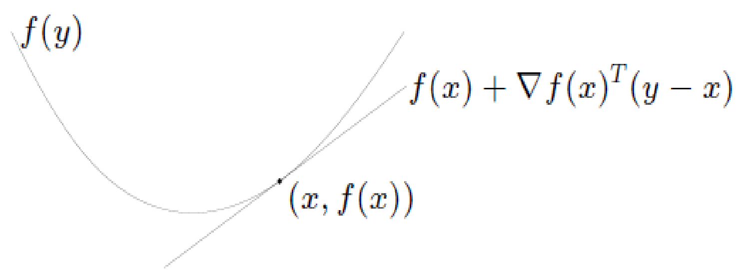
Convex Functions (verification tricks)

- $f(x)$: $R^n \to R$ is convex if and only if the following function of one variable is convex in t for any $x \in dom(f)$ & $v \in \mathbb{R}^n$:

  $$g(t): R \to R: g(t) = f(x + tv), \mathrm{dom(g)} = \{\mathrm{t} \mid \mathrm{x} + \mathrm{tv} \in dom(f)\}$$

- First order condition: Differentiable f with convex domain is convex if and only if

  $$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$



- Second order condition: twice differentiable function f with convex domain is convex if and only if

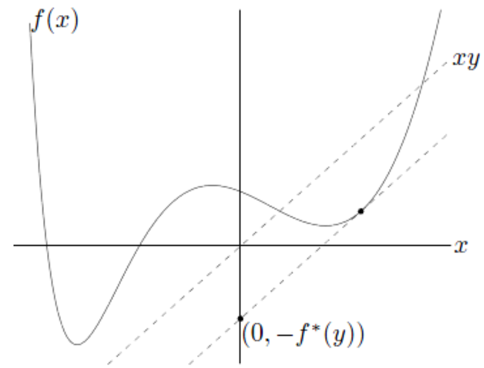
  $$\nabla^2 f(x) \succcurlyeq 0 \; for \; all \; x \; \in dom(f)$$

  - Example: quadratic function $1/2 x^T P x + q^T x + r$ is convex if P is PSD

Operations that preserve convexity:

- Nonnegative weighted sum
  - $\sum_{i=1}^{n} \alpha_i f_i(x)$ is convex if $f_i(x), i = 1,2,..n$ are convex
  - Jensen's inequality: $f(\mathbb{E}(x)) \leq \mathbb{E}f(x)$
- Composition with affine function
  - $f(Ax + b)$ is convex if f(x) is convex
  - Examples: $f(x) = -\sum_{i=1}^{n} \log(b_i - {a_i}^T x)$
- Minimization
  - $g(x) = \min_{y \in C} f(x, y)$ is convex if $f(x, y)$ is convex in (x, y) and C is a convex set
  - Examples: $dist(x, S) = \min_{y \in S} \|x - y\|$ is convex if S is convex
- Perspective $g(x, t) = tf\left(\frac{x}{t}\right), t > 0$
  - Example: $g(x, t) = \frac{x^T x}{t}, t > 0$
- Pointwise maximum and suprimum
  - Piecewise linear function: $f(x) = \max_{i=1,\dots,n} {a_i}^T x + b_i$
  - $g(x) = \sup_{y \in A} f(x, y)$ is convex if $f(x, y)$ is convex in x for each $y \in A$
  - Example: max eigenvalue of a symmetric function $\lambda_{max}(X) = \sup_{\|y\|=1} y^T X y$

Conjugate function:



- The conjugate function of f is defined as

$$f^*(y) = \sup_{x \in dom(f)} (y^T x - f(x))$$

- The conjugate function of f^* is the max cap between the linear function $y^T x$ and $f(x)$. For differentiable functions, this occurs at a point x where $y = \nabla f(x)$

- f^* is convex even if f is not. Because it is a pointwise suprimum of a family of affine functions

- Also known as Lengendre-Fenchel Transformation or Fenchel Transformation

- Examples
  - $f(x) = -\log(x) \rightarrow f^*(y) = -1 - \log(-y)$, $y < 0$
  - $f(x) = \exp(x) \rightarrow f^*(y) = y\log(y) - y$, $y > 0$
  - $f(x) = x\log(x) \rightarrow f^*(y) = \exp(y-1)$, $y \neq 0$
  - $f(x) = 1/x \rightarrow f^*(y) = -2(-y)^{1/2}$, $y \leq 0$

Slack variables

- Converting inequality constraints to equality constrains

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f_0(x) & \rightarrow \\ s.t. & f_i(x) \leq 0 \;\; i = 1, \dots m & \end{array} \qquad \begin{array}{ll} \underset{x, b_i}{\text{minimize}} & f_0(x) \\ s.t. & f_i(x) + b_i = 0 \;\; i = 1, \dots m \\ & b_i \geq 0 \;\; i = 1, \dots m \end{array}$$

- Introducing equality constraints

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f_0(A_0 x + b_0) & \rightarrow \\ s.t. & f_i(A_i x + b_i) \leq 0 \;\; i = 1, \dots m & \end{array} \qquad \begin{array}{ll} \underset{x, b_i}{\text{minimize}} & f_0(y_0) \\ s.t. & f_i(y_i) \leq 0 \;\; i = 1, \dots m \\ & A_i x + b_i = y_i \quad i = 0, \dots m \end{array}$$

- Converting an infeasible problem to feasible by relaxing the constraints

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f_0(x) & \rightarrow \\ s.t. & f_i(x) \leq 0 \;\; i = 1, \dots m & \end{array} \qquad \begin{array}{ll} \underset{x, b_i}{\text{minimize}} & f_0(x) + C \sum_{i=1}^{m} b_i \\ s.t. & f_i(x) - b_i \leq 0 \;\; i = 1, \dots m \\ & b_i \geq 0 \;\; i = 1, \dots m \end{array}$$

Duality

- The following optimization problem

$$\underset{x}{\text{minimize}} \quad f_0(x)$$
$$\textit{subject to} \quad f_i(x) \le 0 \;\; i = 1, \dots m$$
$$h_i(x) = 0 \;\; i = 1, \dots, p$$

- Can be written in the Lagrangian form

$$\mathrm{L}(\mathrm{x}, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- $\lambda_i, i = 1, \dots, m$ are called the Lagrange multipliers associated with the inequalities and $\nu_i, i = 1, \dots, m$ are called the Lagrange multipliers associated with the equalities. They are also called the dual variables.

- The Lagrange dual function is defined as

$$\mathrm{g}(\lambda, \nu) = \inf_x \mathrm{L}(\mathrm{x}, \lambda, \nu) = \inf_x f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- $\mathrm{g}(\lambda, \nu)$ is the lower bound for the optimal value of original problem
  - $\mathrm{g}(\lambda, \nu) \le P^*$

The dual problem

- The following optimization problem is called the dual problem (original problem is called primal)

$$\underset{\lambda,\nu}{\text{maximize}} \;\; \mathrm{g}(\lambda,\nu)$$
$$\mathit{subject\ to}\;\; \lambda \succcurlyeq 0$$

- Finds the best lower bound on p^*
  - A convex optimization problem with optimal value denoted by d^*
  - $\mathrm{L}(\lambda,\nu)$ is concave since it is pointwise infimum of a family of affine functions

$$\mathrm{g}(\lambda,\nu) = \inf_x \mathrm{L}(\mathrm{x},\lambda,\nu) = \inf_x f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

  - This automatically gives a procedure to optimize the non-convex problems.

Solving dual problems

- Solve the dual problem which is convex
- Question: how good it is?
  - The duality gap $p^* - d^*$ is a measure of how good it is
  - Not usually easy to show that the gap is small

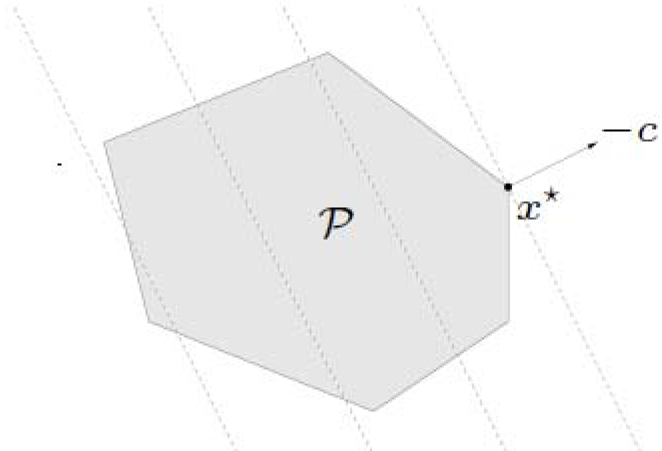
- Strong duality $p^* - d^*$=0
  - Usually (but not always) holds for convex problems
  - Non-convex problem can have strong duality as well so you can get lucky if you use the dual

- If the strong duality holds and x, λ, ν are optimal, then they must satisfy the following conditions, called KKT conditions
  - Primal constraints: $f_i(x) \leq 0,\ i = 1, \ldots m$
  - Dual constraints: $\lambda_i > 0,\ i = 1, \ldots m$
  - Complementary slackness: $\lambda_i f_i(x)$=0 , $i = 1, \ldots m$
  - Gradient of Lagrangian vanishes: $\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x)$

Linear Program (LP)

- Convex problem with affine objective and constraints functions

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x + d \\ s.t. \quad & Gx \leq h \\ & Ax = b \end{aligned}$$

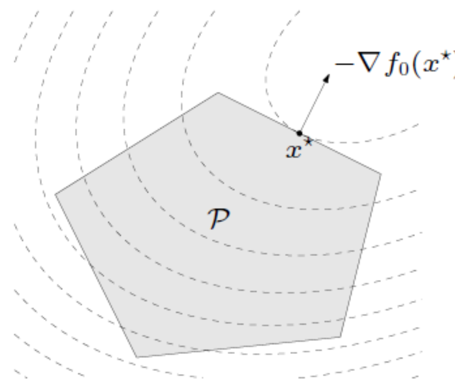


- Feasible set is a polyhedron
- linprog command in MATLAB

Quadratic Program (QP)

- Convex problem with quadratic convex objective and affine constraints functions (P is PSD)

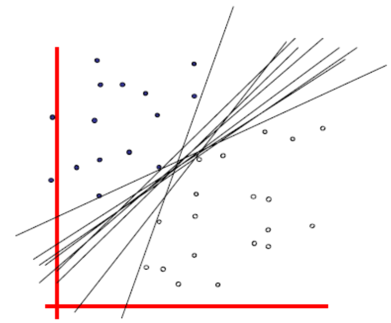
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & 1/2x^T Px + q^T x + r \\ s.t. \quad & Gx \le h \\ & Ax = b \end{aligned}$$



- Minimizes a convex quadratic over a polyhedron
- Quadprog command in matlab

SVM: a QP Example

- Many linear classifiers separating two separable set of examples
- Pick the one with maximum margin



$$\underset{w,b}{\text{minimize}} \quad \|w\|^2$$
$$subject\ to \quad y_i(w^T x_i + b) \geq 1,\ \ i = 1, \dots n$$

- If the examples are not separable, the feasible set of this problem is empty (infeasible problem)
- Utilizing slack variables to relax the constraints and make a feasible problem

$$\underset{w,b}{\text{minimize}} \quad \|w\|^2 + C \sum_{i=1}^{n} \varepsilon_i$$
$$subject\ to \quad y_i(w^T x_i + b) \geq 1 - \varepsilon_i \ \ i = 1, \dots n$$
$$\varepsilon_i \geq 0 \ \ i = 1, \dots, n$$

SVM: dual formulation

- Define the Lagrangian:

$$L(\mathrm{w}, \mathrm{b}, \lambda, \nu) = \quad \|w\|^2 + C \sum\nolimits_{i=1}^{n} \varepsilon_i - \sum_{i=1}^{m} \alpha_i \left(y_i(w^T x_i + b) - 1 + \varepsilon_i\right) - \sum_{i=1}^{n} \mu_i \varepsilon_i$$

- Finding $L(\lambda, \nu) = \inf_{w,b} \mathrm{L}(\mathrm{w}, \mathrm{b}, \lambda, \nu)$

$$\frac{\partial \mathrm{L}(\mathrm{w}, \mathrm{b}, \lambda, \nu)}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i\, y_i x_i$$

$$\frac{\partial \mathrm{L}(\mathrm{w}, \mathrm{b}, \lambda, \nu)}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i\, y_i = 0$$

$$\frac{\partial \mathrm{L}(\mathrm{w}, \mathrm{b}, \lambda, \nu)}{\partial \varepsilon_i} = 0 \rightarrow \alpha_i = C - \mu_i$$

- KKT conditions: 1) $\alpha_i \geq 0,$ 2) $y_i(w^T x_i + b) - 1 + \varepsilon_i \geq 0,$ 3) $\sum_{i=1}^{m} \alpha_i \left(y_i(w^T x_i + b) - 1 + \varepsilon_i\right) = 0,$ 4) $\mu_i \geq 0,$ 5) $\varepsilon_i \geq 0,$ 6) $\mu_i \varepsilon_i = 0$

SVM: dual formulation

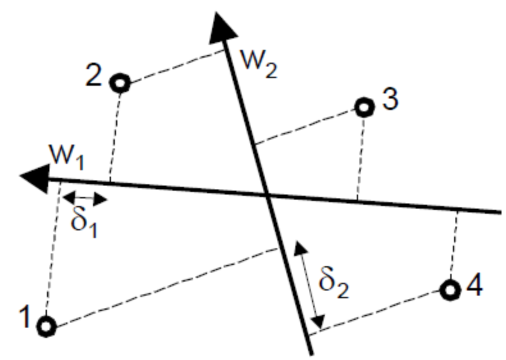
- Using these results, we obtain the dual problem

$$\underset{\lambda,\nu}{\text{maximize}} \ \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j$$
$$subject\ to\ \ 0 \le \alpha_i \le C$$

- Useful form for using the kernel trick

$$\underset{\lambda,\nu}{\text{maximize}} \ \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$subject\ to\ \ 0 \le \alpha_i \le C$$

SVM^Rank: a QP Example



- Ranking problem:
  - n queries $q_i, i = 1,..,n$
  - for query q_i, a list of items ${d_j}^i, j = 1,..,m_i$ (feature vector) with their respect relevancy ${r_j}^i, j = 1,..,m_i$ to the query.
  - Assume also that ${r_j}^i$ are discrete $[1..k]$

- Objective: obtain a linear classifier that respects ordering information
  - Suppose W is such a classifier
  - Construct a set on pair of examples $S = \{(x,z) |\; x = {d_j}^i, z = {d_k}^i, {r_k}^i - {r_j}^i = 1\}$
  - Find W that maximizes the margin between each two items

$$\underset{w,b}{\text{minimize}} \quad \|w\|^2 + C \sum_{r_i - r_j = 1} \varepsilon_{ij}$$

$$\textit{subject to} \quad w^T(x_j - x_i) \geq 1 - \varepsilon_{ij}, \quad (x_i, x_j) \in S$$

$$\varepsilon_{ij} \geq 0 \quad i = 1, \dots, n$$

Multi-Task Learning

- Problem setup
  - T classification problems, each with different set of training examples.
  - Task t has n_t training examples $({x_i}^t, {y_i}^t), i = 1, \ldots, n_t$
  - Feature vector of all task are in the same space
  - Tasks are related (digits recognition, medical domains, etc)

- Objective: to learn linear classifiers $w^t, t = 1, \ldots, T$ for tasks by considering that the tasks are similar

- Solution: assume all tasks are similar to a central unknown task μ

$$\underset{w,b,\mu}{\text{minimize}} \quad \sum_{t=1}^{T} \|w^t\|^2 + \sum_{t=1}^{T} \|w^t - \mu\|^2 + C \sum_{t=1}^{T} \sum_{i=1}^{n} {\varepsilon_i}^t$$

$$subject\ to\ \ y_i \left({w^t}^T {x_i}^t + b^t \right) \geq 1 - {\varepsilon_i}^t, i = 1, \ldots n, t = 1, \ldots, T$$

$${\varepsilon_i}^t \geq 0 \ \ i = 1, \ldots, n, t = 1, \ldots, T$$

- How to write the dual of this problem? (Next lecture)

Quadratically Constrained QP (QCQP)

- Convex problem with quadratic convex objective and constraints functions (P_i are SDP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & 1/2x^T P_0 x + q_0{}^T x + r_0 \\ s.t. \quad & 1/2x^T P_i x + q_i{}^T x + r_i \leq 0 \\ & Ax = b \end{aligned}$$

- Objective and constrains are convex quadratic
- Can be solved with standard toolbox

Semidefinite Programming

- Convex problem with quadratic convex objective and constraints functions

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x + d \\ s.t. \quad & x_1 P_1 + \cdots + x_n P_n + Q \preccurlyeq 0 \text{ (Linear Matrix Inequality)} \\ & Gx \preccurlyeq b \text{ (General inequalities)} \\ & Ax = b \end{aligned}$$

- Or

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & tr(CX) \\ s.t. \quad & tr(A_i X) = b_i \\ & X \succcurlyeq 0 \end{aligned}$$

- If $P_1, \dots, P_n$ *and* Q are all diagonal, the SDP programming reduces to linear programming
- SeDuMi is a good tool to model this type of problems

Local and Global Consistency SSL

- Local and global Consistency, minimize

$$Q(F) = \underbrace{\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} \left\| \frac{F_i}{\sqrt{D_{ii}}} - \frac{F_j}{\sqrt{D_{jj}}} \right\|^2}_{Smoothness} + \underbrace{\mu \sum_{i=1}^{N} \left\| F_i - Y_i \right\|^2}_{Fitting}$$

- Question: convex or non-convex?

$$Q(F) = \quad F \; D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \; F \quad + \underbrace{\mu \sum_{i=1}^{N} \left\| F_i - Y_i \right\|^2}_{Fitting}$$

- How to solve such problems? (Next lecture)

PGDM metric learning

- PGDM metric learning

$$\underset{P}{\text{minimize}} \quad \textstyle\sum_{(x_i, x_j)\in S} \left\| x_i - x_j \right\|_P$$
$$subject\ to \ \textstyle\sum_{(x_i, x_j)\in D} \left\| x_i - x_j \right\|_P \geq 1$$
$$P \succcurlyeq 0$$

- Question: convex or non-convex?
- How should we solve such problems? (next lecture)

LMNN metric learning

- LMNN metric learning

$$\underset{A}{\text{minimize}} \quad \sum_{(x_i, x_j) \in S} \left\| x_i - x_j \right\|_P$$
$$s.t \ \ \|x_i - x_k\|_P - \left\|x_i - x_j\right\|_P \geq 1, (x_i,\ x_j,\ x_k) \in R$$
$$P \succcurlyeq 0$$

- $in\ (x_i,\ x_j,\ x_k) \in R$, $\left(x_i,\ x_j\right)$ are of the same class and neighbor according to Euclidean distance. $(x_i,\ x_k)$ are from two different classes.
- Question: convex or non-convex?
- How should we solve such problems? (next lecture)