EE269
Signal Processing for Machine Learning

Lecture 10

Instructor : Mert Pilanci

Stanford University

February 11, 2019

Recap: Separating Hyperplanes

- Linear/Quadratic Discriminant Analysis, Bayes Optimal Classifiers

  require modeling signals: $p(x)$, and class distributions $p(y|x)$
- **Vapnik's Principle**

  'When solving a problem if interest, do not solve a more general problem as an intermediate step'

Recap: Separating Hyperplanes

- Directly estimate hyperplanes $w^T x + b \geq 0$
  parameters $\theta = (w, b)$

Recap: Separating Hyperplanes

- Directly estimate hyperplanes $w^T x + b \geq 0$
  parameters $\theta = (w, b)$
- Hyperplane: $H = \left\{x : \; w^T x + b = 0\right\}$

Recap: Separating Hyperplanes

- Directly estimate hyperplanes $w^T x + b \geq 0$
  parameters $\theta = (w, b)$
- Hyperplane: $H = \left\{x : \; w^T x + b = 0\right\}$
  distance between a point z and H
  $d(z, H) = \min_{h \in H} ||z - h||_2$
  Decompose $z = z_0 + \frac{w}{||w||_2} r$
  $w^T z + b = w^T z_0 + b + ||w||_2 r$

Recap: Separating Hyperplanes

- Directly estimate hyperplanes $w^T x + b \geq 0$
  parameters $\theta = (w, b)$
- Hyperplane: $H = \left\{x : \; w^T x + b = 0\right\}$
  distance between a point z and H
  $d(z, H) = \min_{h \in H} ||z - h||_2$
  Decompose $z = z_0 + \frac{w}{||w||_2} r$
  $w^T z + b = w^T z_0 + b + ||w||_2 r$

$$d(z, H) = |r| = \frac{|w^T z + b|}{||w||_2}$$

Margin

Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

- Directly estimate hyperplanes $w^T x + b = 0$
- Margin ρ of a hyperplane is

$$\begin{aligned} \rho(w, b) &= \min_{i=1,...,n} d(x_i, H) \\ &= \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2} \end{aligned}$$

Margin

Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

- Directly estimate hyperplanes $w^T x + b = 0$
- Margin ρ of a hyperplane is

$$\begin{aligned}\rho(w, b) &= \min_{i=1,...,n} d(x_i, H) \\ &= \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2}\end{aligned}$$

- Maximum margin separating hyperplane is the solution of

$$\begin{aligned}&\max_{w,b} \rho(w, b) \\ &s.t. \;\; y_i(w^T x_i + b) \geq 0 \; \forall i\end{aligned}$$

Maximum margin hyperplane

Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

- Margin $\rho(w, b) = \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2}$
- Maximum margin separating hyperplane is the solution of

$$\max_{w,b} \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2}$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 0 \; \forall i$$

Maximum margin hyperplane

Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

- Margin $\rho(w, b) = \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2}$
- Maximum margin separating hyperplane is the solution of

$$\max_{w,b} \min_{i=1,...,n} \frac{|w^T x_i + b|}{||w||_2}$$
$$s.t. \quad y_i(w^T x_i + b) \geq 0 \ \forall i$$

- **not unique**

  $(\alpha w, \alpha b)$ gives the same hyperplane

Maximum margin hyperplane

Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

- Margin $\rho(w, b) = \min_{i=1,\ldots,n} \frac{|w^T x_i + b|}{||w||_2}$
- Maximum margin separating hyperplane is the solution of

$$\max_{w,b} \min_{i=1,\ldots,n} \frac{|w^T x_i + b|}{||w||_2}$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 0 \; \forall i$$

- **not unique**

  $(\alpha w, \alpha b)$ gives the same hyperplane
- Scale w and b by $\frac{1}{\min_{i=1,\ldots,n} |w^T x_i + b|}$

  Now $\rho = \frac{1}{||w||_2}$

Maximum margin hyperplane

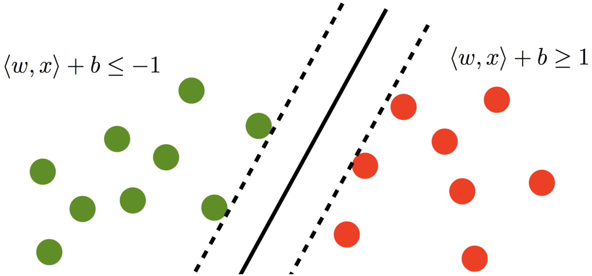
Data $x_1, ..., x_n$ and corresponding labels $y_1, ...y_n \in \{-1, +1\}$

$$\max_{w,b} \frac{1}{||w||_2}$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 1 \;\forall i$$

equivalently

$$\min_{w,b} ||w||_2^2$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 1 \;\forall i$$

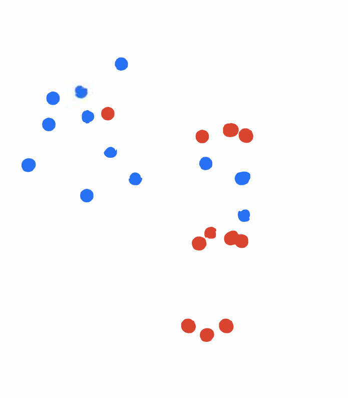
- Hard-margin support vector machine (SVM)



slide credit: Alex Smola

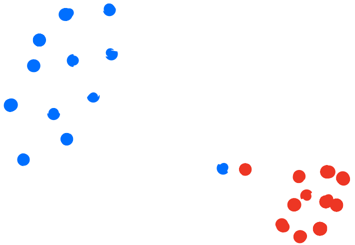
Problems with hard margin

- Separability



Problems with hard margin

- Sensitivity



Soft Margin Support Vector Machine

$$\min_{w,b,s_1,\dots,s_n} \frac{1}{2}||w||_2^2 + C\frac{1}{n}\sum_{i=1}^{n} s_i$$

$$\begin{aligned} s.t.\quad & y_i(w^T x_i + b) \geq 1 - s_i \ \forall i \\ & s_i \geq 0 \ \forall i \end{aligned}$$

- $s_1, \dots, s_n$ are slack variables
- C is a tuning parameter

Constrained Optimization

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to } & g_i(x) \leq 0 \;\; \forall i \in 1, ..., m \\ & h_j(x) = 0 \;\; \forall j \in 1, ..., n \end{aligned}$$

- optimization variable x is feasible if it satisfies all the constraints

Constrained Optimization and Lagrangian

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{subject to } & g_i(x) \leq 0 \quad \forall i \in 1, ..., m \\
& h_j(x) = 0 \quad \forall j \in 1, ..., n
\end{aligned}
$$

- Lagrangian:

$$
\begin{aligned}
\mathcal{L}(x, \lambda, \mu) &= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m} \mu_j h_j(x) \\
&= f(x) + \lambda^T g(x) + \mu^T h(x)
\end{aligned}
$$

where $\lambda_i \geq 0 \ \forall i$

Dual Function

$$\min_{x} \quad f(x)$$
$$\text{subject to } g_i(x) \leq 0 \quad \forall i \in 1, ..., m$$
$$h_j(x) = 0 \quad \forall j \in 1, ..., n$$

- Lagrangian:

$$\begin{aligned}\mathcal{L}(x, \lambda, \mu) &= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m} \mu_j h_j(x) \\ &= f(x) + \lambda^T g(x) + \mu^T h(x)\end{aligned}$$

- dual function:

$$F(\lambda, \mu) = \inf_{x} \mathcal{L}(x, \lambda, \mu)$$

- infimum (inf) : Greatest lower bound

Dual Problem

$$
\begin{aligned}
p^* := \min_{x} \quad & f(x) \\
\text{subject to } & g_i(x) \leq 0 \quad \forall i \in 1, ..., m \\
& h_j(x) = 0 \quad \forall j \in 1, ..., n
\end{aligned}
$$

- Lagrangian:

$$
\begin{aligned}
\mathcal{L}(x, \lambda) &= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m} \mu_j h_j(x) \\
&= f(x) + \lambda^T g(x) + \mu^T h(x)
\end{aligned}
$$

- dual function:

$$F(\lambda, \mu) = \inf_{x} \mathcal{L}(x, \lambda)$$

- dual problem:

$$d^* = \max F(\lambda)$$

Dual Problem

$$
\begin{aligned}
p^* := \min_x \quad & f(x) \\
\text{subject to } & g_i(x) \leq 0 \quad \forall i \in 1, ..., m \\
& h_j(x) = 0 \quad \forall j \in 1, ..., n
\end{aligned}
$$

- Lagrangian:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)$$

- dual function:

$$F(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$$

- dual problem: $d^* = \max_{\lambda \geq 0, \mu} F(\lambda, \mu)$
- weak duality: $F(\lambda, \mu) \leq f(x^*)$, where p^* solves the primal problem, i.e., $d^* \leq p^*$

Dual Problem

$$
\begin{aligned}
p^* := \min_x \quad & f(x) \\
\text{subject to } & g_i(x) \leq 0 \;\; \forall i \in 1, ..., m \\
& h_j(x) = 0 \;\; \forall j \in 1, ..., n
\end{aligned}
$$

- Lagrangian:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)$$

- dual function:

$$F(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$$

- dual problem: $d^* = \max_{\lambda \geq 0, \mu} F(\lambda, \mu)$

Dual Problem

- Lagrangian

$$\mathcal{L}(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)$$

- primal problem:

$$\min_{x} \max_{\lambda_i \geq 0, \mu_j \forall i,j} L(x,\lambda,\mu)$$

- dual problem:

$$\max_{\lambda_i \geq 0, \mu_j \forall i,j} \min_{x} L(x,\lambda,\mu)$$

Dual Problem

- Lagrangian

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x)$$

- primal problem:

$$\min_{x} \max_{\lambda_i \geq 0, \mu_j \forall i,j} L(x, \lambda, \mu)$$

- dual problem:

$$\max_{\lambda_i \geq 0, \mu_j \forall i,j} \min_{x} L(x, \lambda, \mu)$$

- weak duality:

$$\min_{x} \max_{\lambda_i \geq 0, \mu_j \forall i,j} L(x, \lambda, \mu) \geq \max_{\lambda_i \geq 0, \mu_j \forall i,j} \min_{x} L(x, \lambda, \mu)$$

Strong Duality

- If $p^* = d^*$, we say **strong duality** holds

  Theorem: If $f, g_1, ..., g_m$ are convex, $h_1, ..., h_n$ are affine, and a constraint qualification holds, then $p^* = d^*$

Strong Duality

- If $p^* = d^*$, we say **strong duality** holds

  Theorem: If $f, g_1, ..., g_m$ are convex, $h_1, ..., h_n$ are affine, and a constraint qualification holds, then $p^* = d^*$
- Examples of constraint qualification

  Slater's condition: $\exists x$ such that $g_i(x) < 0$ and $h_j(x) = 0 \; \forall i, j$

Strong Duality

- If $p^* = d^*$, we say **strong duality** holds

  Theorem: If $f, g_1, ..., g_m$ are convex, $h_1, ..., h_n$ are affine, and a constraint qualification holds, then $p^* = d^*$
- Examples of constraint qualification

  Slater's condition: $\exists x$ such that $g_i(x) < 0$ and $h_j(x) = 0 \ \forall i, j$
- example: $0 \leq x_i \leq 1, x \in \mathbb{R}^d$, Slater's condition holds

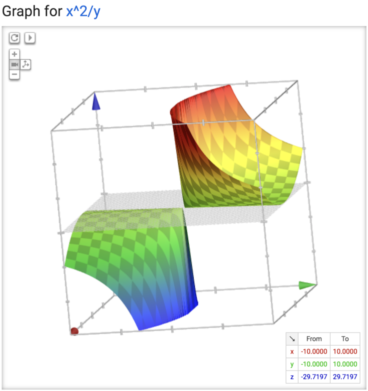
Strong Duality

- If $p^* = d^*$, we say **strong duality** holds

  Theorem: If $f, g_1, ..., g_m$ are convex, $h_1, ..., h_n$ are affine, and a constraint qualification holds, then $p^* = d^*$
- Examples of constraint qualification

  Slater's condition: $\exists x$ such that $g_i(x) < 0$ and $h_j(x) = 0 \ \forall i, j$
- example: $0 \leq x_i \leq 1, x \in \mathbb{R}^d$, Slater's condition holds
- example: $\frac{x_1^2}{x_2} \leq 0, \ x_1 \in \mathbb{R}, x_2 \in \mathbb{R}_+$

Slater's Condition Fails



- Exercise: show that strong duality doesn't hold:

$$\min_{\frac{x_1^2}{x_2} \le 0,\ x_1 \in \mathbb{R}, x_2 \in \mathbb{R}_+} \quad e^{-x_1}$$

KKT conditions

- **Theorem** If $p^* = d^*$, x^* is promal optimal, (λ^*, μ^*) dual optimal, then Karush-Kuhn-Tucker (KKT) hold
  1. Stationarity $\nabla_x L(x, \lambda, \mu) = 0$ at (x^*, λ^*, μ^*)
  2. Primal feasibility $g_i(x^*) \leq 0$, $h_j(x^*) = 0 \ \forall i, j$
  3. Dual feasibility $\lambda_i^* \geq 0$, $\forall i$
  4. Complementary slackness: $\lambda_i^* g_i(x^*) = 0$, $\forall i$

KKT conditions

- **Theorem** If $p^* = d^*$, x^* is promal optimal, (λ^*, μ^*) dual optimal, then Karush-Kuhn-Tucker (KKT) hold
  1. Stationarity $\nabla_x L(x, \lambda, \mu) = 0$ at (x^*, λ^*, μ^*)
  2. Primal feasibility $g_i(x^*) \leq 0$, $\; h_j(x^*) = 0 \; \forall i, j$
  3. Dual feasibility $\lambda_i^* \geq 0$, $\forall i$
  4. Complementary slackness: $\lambda_i^* g_i(x^*) = 0$, $\; \forall i$
- **Theorem** If $f, g_1, ..., g_m$ convex, $h_1, ..., h_m$ affine and (x^*, λ^*, μ^*) satisfy KKT, then (x^*, λ^*, μ^*) is optimal and strong duality holds.

Soft Margin Support Vector Machine

$$
\begin{aligned}
&\min_{w,b,s_1,\dots,s_n} \ \frac{1}{2}||w||_2^2 + C\frac{1}{n}\sum_{i=1}^{n} s_i \\
s.t. \ \ & y_i(w^T x_i + b) \geq 1 - s_i \ \forall i \\
& s_i \geq 0 \ \forall i
\end{aligned}
$$

- Lagrangian

$$
\begin{aligned}
L(w,b,s_i,\alpha,\beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i (y_i(w^T x_i + b) + s_i - 1) \\
- \sum_i \beta_i s_i
\end{aligned}
$$

Soft Margin Support Vector Machine

$$\min_{w,b,s_1,\ldots,s_n} \frac{1}{2}||w||_2^2 + C\frac{1}{n}\sum_{i=1}^{n} s_i$$
$$s.t. \quad y_i(w^T x_i + b) \geq 1 - s_i \ \forall i$$
$$s_i \geq 0 \ \forall i$$

- Lagrangian

$$L(w,b,s_i,\alpha,\beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1) - \sum_i \beta_i s_i$$

- dual problem $\max_{\alpha \geq 0, \beta_0} F(\alpha, \beta)$
  $F(\alpha,\beta) = \min_{w,b,s_i} L(w,b,s_i,\alpha,\beta)$

Lagrangian

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1)$$
$$- \sum_i \beta_i s_i$$

- $\frac{\partial L}{\partial w} = 0 = w - \sum_i \alpha_i y_i x_i, \quad \implies w^* = \sum_i \alpha_i y_i x_i$

Lagrangian

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1) - \sum_i \beta_i s_i$$

- $\frac{\partial L}{\partial w} = 0 = w - \sum_i \alpha_i y_i x_i, \quad \implies w^* = \sum_i \alpha_i y_i x_i$
- $\frac{\partial L}{\partial b} = 0 = \sum_i \alpha_i y_i = 0$

Lagrangian

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1) - \sum_i \beta_i s_i$$

- $\frac{\partial L}{\partial w} = 0 = w - \sum_i \alpha_i y_i x_i, \quad \implies w^* = \sum_i \alpha_i y_i x_i$
- $\frac{\partial L}{\partial b} = 0 = \sum_i \alpha_i y_i = 0$
- $\frac{\partial L}{\partial s_i} = 0 = \frac{C}{n} - \alpha_i - \beta_i$

Lagrangian

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1)$$
$$- \sum_i \beta_i s_i$$

- $\frac{\partial L}{\partial w} = 0 = w - \sum_i \alpha_i y_i x_i, \quad \Longrightarrow \; w^* = \sum_i \alpha_i y_i x_i$
- $\frac{\partial L}{\partial b} = 0 = \sum_i \alpha_i y_i = 0$
- $\frac{\partial L}{\partial s_i} = 0 = \frac{C}{n} - \alpha_i - \beta_i$
- $F(\alpha, \beta) = -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$

Lagrangian and Dual SVM

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i(y_i(w^T x_i + b) + s_i - 1)$$
$$- \sum_i \beta_i s_i$$

- $F(\alpha, \beta) = -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$
- dual problem

$$\max_{\alpha,\beta} -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$
$$\text{subject to } \sum_i \alpha_i y_i = 0$$
$$\alpha_i + \beta_i = \frac{C}{n}$$
$$\alpha_i \geq 0, \beta_i \geq 0$$

Lagrangian and Dual SVM

$$L(w, b, s_i, \alpha, \beta) = \frac{1}{2}||w||_2^2 + \frac{C}{n}\sum_{i=1}^{n} s_i - \sum_i \alpha_i (y_i(w^T x_i + b) + s_i - 1) - \sum_i \beta_i s_i$$

- $F(\alpha, \beta) = -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$
- dual problem

$$\max_\alpha -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$

$$\text{subject to } \sum_i \alpha_i y_i = 0$$

$$0 \le \alpha_i \le \frac{C}{n}$$

Primal SVM vs Dual SVM

- primal problem

$$
\min_{w,b,s_1,\ldots,s_n} \ \frac{1}{2}||w||_2^2 + C\frac{1}{n}\sum_{i=1}^{n} s_i
$$

$$
\begin{aligned}
s.t. \ \ & y_i(w^T x_i + b) \geq 1 - s_i \ \forall i \\
& s_i \geq 0 \ \forall i
\end{aligned}
$$

- dual problem

$$
\max_{\alpha} -\frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i
$$

$$
\text{subject to } \sum_i \alpha_i y_i = 0
$$

$$
0 \leq \alpha_i \leq \frac{C}{n}
$$

- $w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i$

Geometry of SVM

- dual problem

$$\max_{\alpha} -\frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i$$
$$\text{subject to } \sum_i \alpha_i y_i = 0$$
$$0 \le \alpha_i \le \frac{C}{n}$$

- $C \to \infty$ gives hard margin SVM

$$\max_{\alpha} -\frac{1}{2}\Big|\Big|\sum_i \alpha_i y_i x_i\Big|\Big|_2^2 + \sum_i \alpha_i$$
$$\text{subject to } \sum_i \alpha_i y_i = 0$$
$$0 \le \alpha_i$$

Geometry of SVM

- dual problem

$$\max_{\alpha} -\frac{1}{2}\Big|\Big|\sum_i \alpha_i y_i x_i\Big|\Big|_2^2 + \sum_i \alpha_i$$
$$\text{subject to } \sum_i \alpha_i y_i = 0$$
$$0 \le \alpha_i$$

- $C \to \infty$ gives hard margin SVM

$$\max_{\alpha} -\frac{1}{2}\Big|\Big|\sum_{i\in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i\Big|\Big|_2^2 + \sum_i \alpha_i$$
$$\text{subject to } \sum_{i\in +} \alpha_i = \sum_{i\in -} \alpha_i$$
$$0 \le \alpha_i$$

Geometry of SVM

$$\max_{\alpha} -\frac{1}{2}\Big|\Big|\sum_{i\in +}\alpha_i x_i - \sum_{i\in -}\alpha_i x_i\Big|\Big|_2^2 + \sum_{i\in +}\alpha_i + \sum_{i\in -}\alpha_i$$

$$\text{subject to } \sum_{i\in +}\alpha_i = \sum_{i\in -}\alpha_i$$

$$0 \leq \alpha_i$$

Geometry of SVM

$$\max_{\alpha} -\frac{1}{2}\Big\|\sum_{i\in+}\alpha_i x_i - \sum_{i\in-}\alpha_i x_i\Big\|_2^2 + \sum_{i\in+}\alpha_i + \sum_{i\in-}\alpha_i$$

$$\text{subject to } \sum_{i\in+}\alpha_i = \sum_{i\in-}\alpha_i$$

$$0 \le \alpha_i$$

- impose the constraint $\sum_{i\in+}\alpha_i = \gamma$ for some $\gamma > 0$ and maximize over γ to obtain the same problem

$$\max_{\gamma\ge 0}\max_{\alpha} -\frac{1}{2}\Big\|\sum_{i\in+}\alpha_i x_i - \sum_{i\in-}\alpha_i x_i\Big\|_2^2 + 2\gamma$$

$$\text{subject to } \sum_{i\in+}\alpha_i = \sum_{i\in-}\alpha_i = \gamma$$

$$0 \le \alpha_i$$

$$
\begin{aligned}
&\max_{\gamma \geq 0} \max_{\alpha} -\frac{1}{2} \Big\| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i \Big\|_2^2 + 2\gamma \\
&\text{subject to } \sum_{i \in +} \alpha_i = \sum_{i \in -} \alpha_i = \gamma \\
&\qquad 0 \leq \alpha_i
\end{aligned}
$$

- variable change $\alpha_i \leftarrow \gamma \alpha_i$

$$
\begin{aligned}
&\max_{\gamma \geq 0} \max_{\alpha} -\frac{\gamma^2}{2} \Big\| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i \Big\|_2^2 + 2\gamma \\
&\text{subject to } \sum_{i \in +} \alpha_i = \sum_{i \in -} \alpha_i = 1 \\
&\qquad 0 \leq \alpha_i
\end{aligned}
$$

$$\max_{\gamma \geq 0} \max_{\alpha} -\frac{\gamma^2}{2} \Big|\Big| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i \Big|\Big|_2^2 + 2\gamma$$

$$\text{subject to } \sum_{i \in +} \alpha_i = \sum_{i \in -} \alpha_i = 1$$

$$0 \leq \alpha_i$$

- optimize over γ: $-\gamma^*(|| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i||_2^2) + 2 = 0$

$$\gamma^* = \frac{2}{|| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i||_2^2}$$

Geometry of SVM

- plug in optimal γ^*

$$\min_{\alpha} \Big|\Big| \sum_{i \in +} \alpha_i x_i - \sum_{i \in -} \alpha_i x_i \Big|\Big|_2^2$$
$$\text{subject to } \sum_{i \in +} \alpha_i = \sum_{i \in -} \alpha_i = 1$$
$$0 \leq \alpha_i$$

- convex hull

Geometry of SVM

- plug in optimal γ^*

$$\min_{\alpha} \left\| \sum_{i\in+} \alpha_i x_i - \sum_{i\in-} \alpha_i x_i \right\|_2^2$$

$$\text{subject to } \sum_{i\in+} \alpha_i = \sum_{i\in-} \alpha_i = 1$$

$$0 \le \alpha_i$$

- minimum distance between convex hulls

$$\min_{\substack{u\in\mathcal{H}(S_1)\\ v\in\mathcal{H}(S_0)}} \|u-v\|^2$$