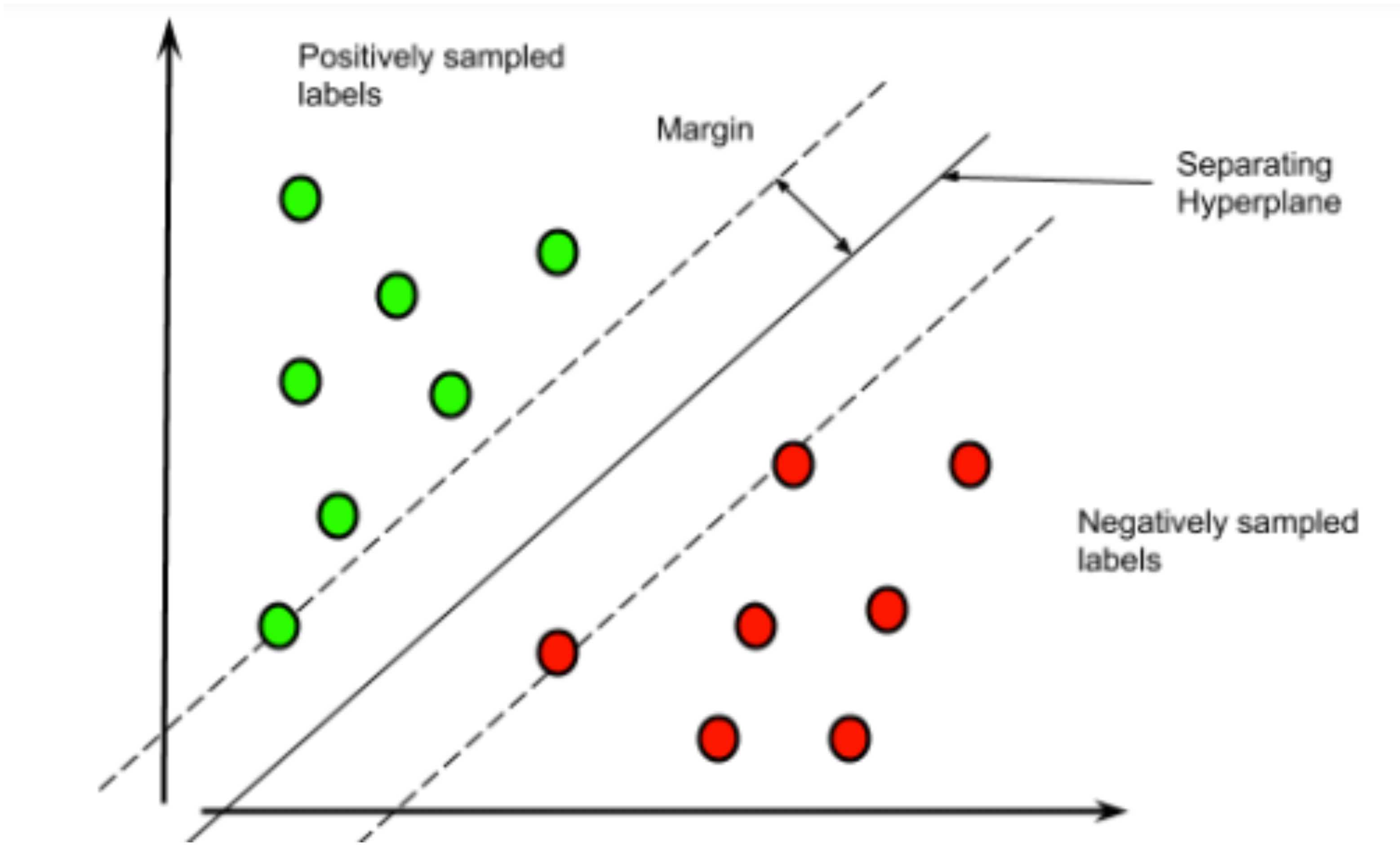



# Lagrange Relaxation, Duality, and the Dual Formulation of Support Vector Machines (SVM)

Leandro L. Minku

# Overview

- Lagrange relaxation and duality
- Dual representation of SVM
- Kernel trick

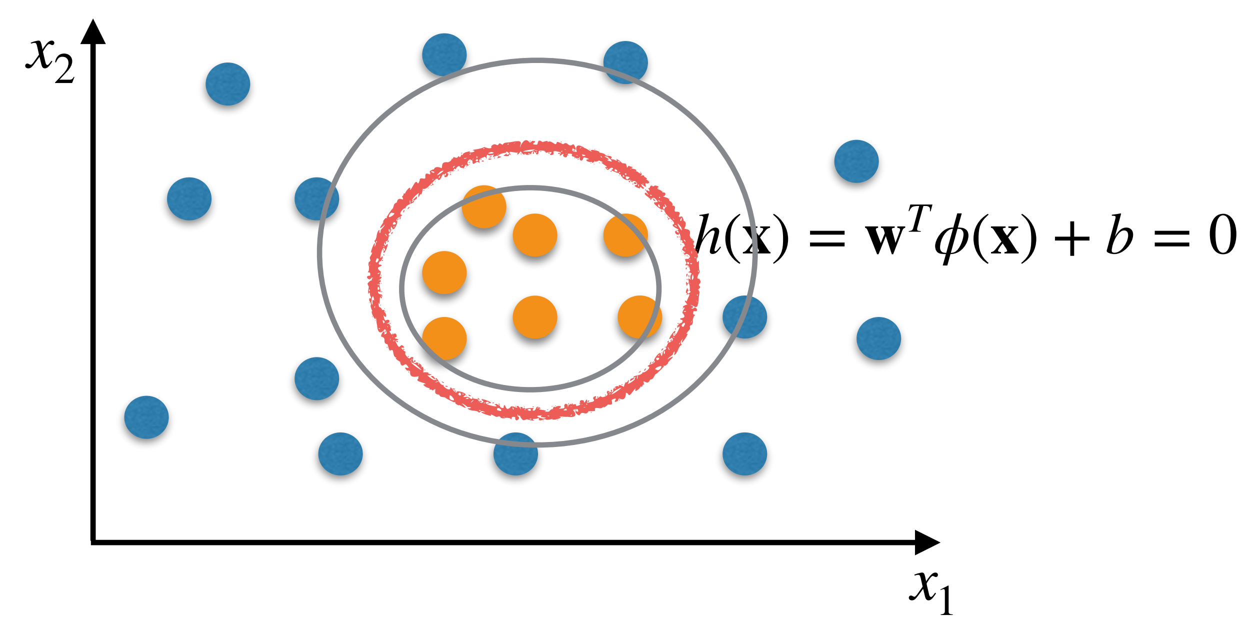
# Maximum Margin Classifiers With Basis Expansion

$$\underset{\mathbf{w},b}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$$

Subject to

$$y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}) + b) \geq 1,$$

$$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}.$$




It is possible to use $\phi(\mathbf{x}) = \mathbf{x}$ if we wish.

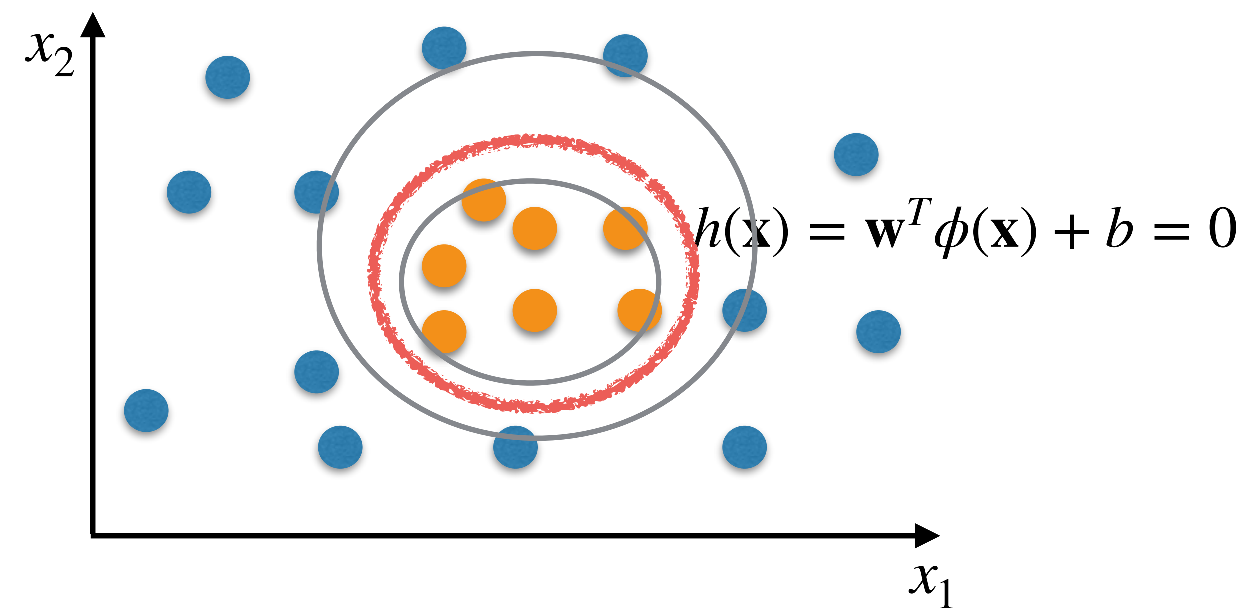
# Maximum Margin Classifiers With Basis Expansion

$$\underset{\mathbf{w},b}{\text{argmin}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$$

Subject to

$$y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}) + b) \geq 1,$$

$$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}.$$




Depending on $\phi(\mathbf{x})$, its computation can be very expensive, as it may be taking us to a very high dimensional problem.

# Rewriting Our Optimisation Problem

Primal Formulation

$$\underset{\mathbf{w},b}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$$

Subject to: $y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) \geq 1$

$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

Dual Formulation

We will get rid of $\mathbf{w}$ and $b$ (!!!!!)

We will get rid of $\mathbf{w}^T \phi(\mathbf{x}^{(n)})$ and $\|\mathbf{w}\|^2$

# Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f(\mathbf{x}) \leq 0$

Lagrange Relaxation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + af(\mathbf{x}) \right\}$$

+

where $a \geq 0$ is called a Lagrange multiplier and
$L(\mathbf{x}, a) = F(\mathbf{x}) + af(\mathbf{x})$ is called the Lagrangian.

We will be penalising the objective when the constraint is violated (unless $a = 0$).

So, we will be searching for a solution that does not violate the constraint.

# Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

Lagrange Relaxation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

We will be penalising the objective when a constraint is violated (unless its $a_i = 0$).

# Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

Lagrange Relaxation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

However, we need to choose appropriate values for each $a_i$ to ensure that the penalty is large enough.

# Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

Lagrange Relaxation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

We will be rewarding the objective when a constraint is not violated (unless its $a_i = 0$).

# Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

Lagrange Relaxation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

When a feasible solution is found, $L(\mathbf{x}, \mathbf{a}) \leq F(\mathbf{x})$.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

When the constraint $f_i(\mathbf{x})$ is violated, this is +

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\} \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

When a constraint is violated, the penalty will become infinitely large.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

When the constraint $f_i(\mathbf{x})$ is violated, this is +

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\} \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

So, an optimisation algorithm will try to avoid violations as much as possible.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0, \ i \in \{1, \cdots, N\}$

When the constraint $f_i(\mathbf{x})$ is not violated, this is either -

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\} \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

0

where $a_i \geq 0, \ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

any value $\geq 0$

or this is 0

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\} \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

When no constraint is violated, the penalty is zero.
So, $L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x})$.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

Minimax Primal Formulation

$$\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\} \quad \min_{\mathbf{x}} \max_{\mathbf{a}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

Problem: needs to solve a constrained optimisation problem (max) inside an unconstrained (min) one.

# Dual Formulation

Minimax Primal Formulation

$$\min_{\mathbf{x}} \max_{\mathbf{a}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

"Dual function"

Dual Formulation

$$\max_{\mathbf{a}} \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

The dual *may* be easier to solve, e.g., it may be that it is possible to solve the unconstrained min problem in closed form, and it may be possible to reduce the number of variables.

However, it is not always equivalent to the primal, which is known as weak duality.

# Strong Duality: Minimax = Maxmin

max Primal
rmulation

$$\min_{\mathbf{x}} \max_{\mathbf{a}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

Convex in $\mathbf{x}$ (this is true when both $F(\mathbf{x})$ and $f_i(\mathbf{x})$ are convex)

Concave in $\mathbf{a}$

and if there is at least one feasible $\mathbf{x}$ satisfying $f_i(\mathbf{x}) < 0,\ \forall i$ (this happens for most ML problems)

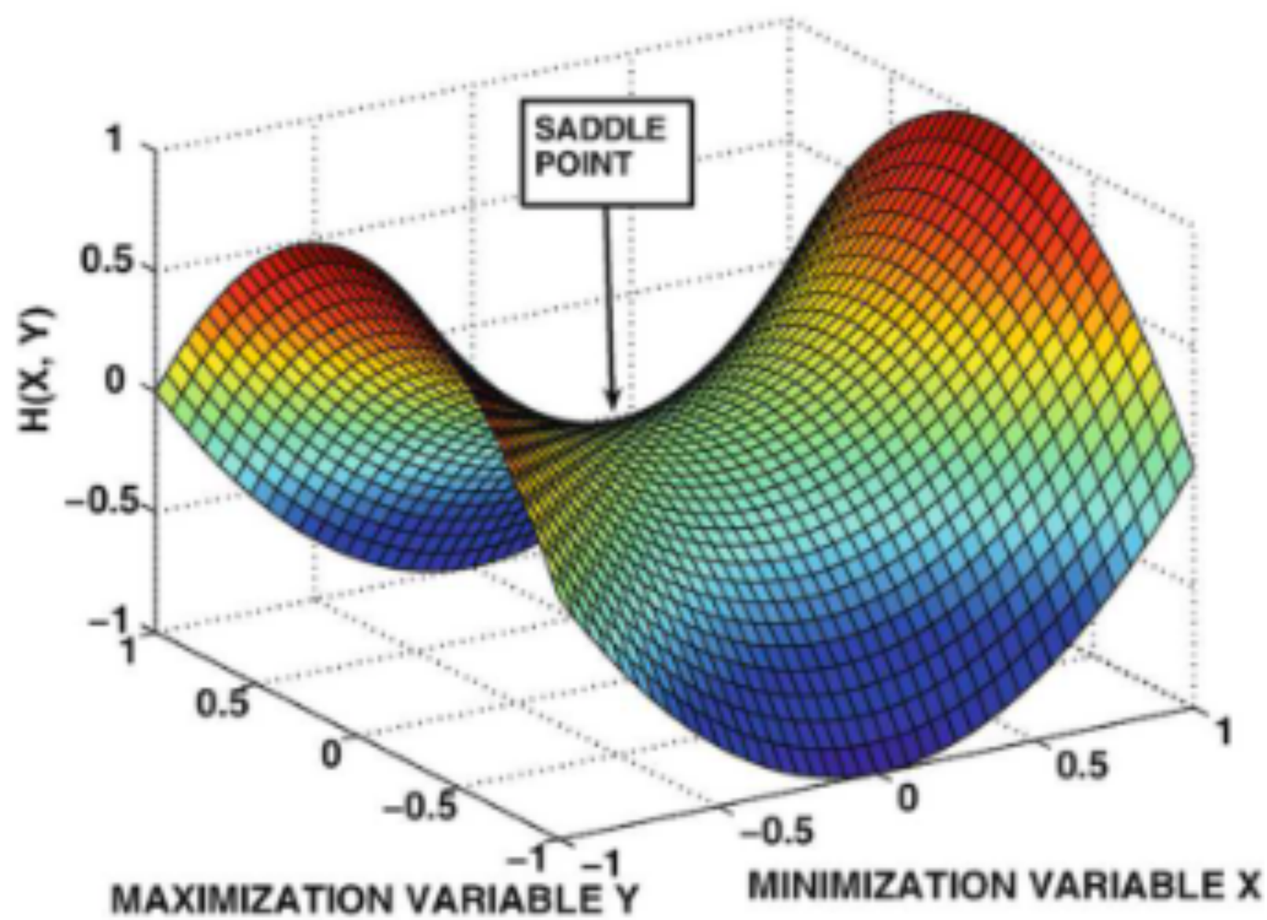
Dual
Formulatio

$$\max_{\mathbf{a}} \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

# Minimax = Maxmin Illustrative Example



# Karush–Kuhn–Tucker (KKT) Conditions

- In unconstrained convex optimisation, a necessary and sufficient condition for optimality is $\nabla F(\mathbf{x}) = 0$.

- KKT are the necessary and sufficient conditions for optimality in a convex optimisation problem $\min_{\mathbf{x}} F(\mathbf{x})$ with convex inequality constraints $f_i(\mathbf{x}) \leq 0$.

- A solution $\mathbf{x}$ is optimal for the primal and a solution $\mathbf{a}$ is optimal for the dual iif:
  - Stationarity: note that in $\max_{\mathbf{a}} \min_{\mathbf{x}}$, for each fixed value of $\mathbf{a}$, we will minimise an unconstrained problem with respect to $\mathbf{x}$. So,

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{a}) = \nabla_{\mathbf{x}} F(\mathbf{x}) + \sum_{i=1}^{N} a_i \nabla_{\mathbf{x}} f_i(\mathbf{x}) = 0$$

  - Complementary slackness: $a_i f_i(\mathbf{x}) = 0, \forall i \in \{1, \cdots, N\}$.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

When the constraint $f_i(\mathbf{x})$ is not violated, this may be -

$$\min_{\mathbf{x}} F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} \underset{0}{\downarrow} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

# Minimax Primal Formulation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, N\}$

any value $\geq 0$

When the constraint $f_i(\mathbf{x})$ is not violated, this may be 0

$$\min_{\mathbf{x}} F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \longrightarrow \min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \max_{\mathbf{a}} \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right\}$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

# Karush–Kuhn–Tucker (KKT) Conditions

- In unconstrained convex optimisation, a necessary and sufficient condition for optimality is $\nabla F(\mathbf{x}) = 0$.
- KKT are the necessary and sufficient conditions for optimality in a convex optimisation problem $\min_{\mathbf{x}} F(\mathbf{x})$ with convex inequality constraints $f_i(\mathbf{x}) \leq 0$.
- A solution $\mathbf{x}$ is optimal for the primal and a solution $\mathbf{a}$ is optimal for the dual iif:
  - Stationarity: note that in $\max_{\mathbf{a}} \min_{\mathbf{x}}$, for each fixed value of $\mathbf{a}$, we will minimise an unconstrained problem with respect to $\mathbf{x}$. So,
  $$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{a}) = \nabla F(\mathbf{x}) + \sum_{i=1}^{N} a_i \nabla f_i(\mathbf{x}) = 0$$
  - Complementary slackness: $a_i f_i(\mathbf{x}) = 0, \forall i \in \{1, \cdots, N\}$.
  - Feasibility: the primal and dual constraints must be satisfied.
    - $f_i(\mathbf{x}) \leq 0, \forall i \in \{1, \cdots, N\}$
    - $a_i \geq 0, \forall i \in \{1, \cdots, N\}$

# SVM: From Primal to Dual

Primal Formulation

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

Subject to: $y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) \geq 1$

$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

$\Downarrow$

Subject to: $y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) - 1 \geq 0$

Subject to: $1 - y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) \leq 0$

Lagrange Relaxation

# Recap: Lagrange Relaxation

Primal Formulation

$$\min_{\mathbf{x}} F(\mathbf{x})$$

Subject to: $f_i(\mathbf{x}) \leq 0,\ i \in \{1, \cdots, n\}$

Lagrange Relaxation

$$\min_{\mathbf{x}} F(\mathbf{x}) + \sum_{i=1}^{n} a_i f_i(\mathbf{x})$$

where $a_i \geq 0,\ i \in \{1, \cdots, n\}$ are the Lagrange multipliers and

$L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{n} a_i f_i(\mathbf{x})$ is the Lagrangian.

# SVM: From Primal to Dual

Primal Formulation

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

Subject to: $y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) \geq 1$

$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

$\Downarrow$

Subject to: $y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) - 1 \geq 0$

Subject to: $1 - y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b) \leq 0$

Lagrange Relaxation

$$\min_{\mathbf{w},b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)} (1 - y^{(n)}(\mathbf{w}^T \phi(\mathbf{x}^{(n)}) + b)) \right\}$$

Where $N$ is the number of training examples.

Subject to: $a^{(n)} \geq 0, \ \forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

# Recap: Minimax Primal and Dual Formulation

Minimax Primal Formulation

$$\min_{\mathbf{x}} \max_{\mathbf{a}} \left( F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right)$$

Dual Formulation

$$\max_{\mathbf{a}} \min_{\mathbf{x}} \left( F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x}) \right)$$

where $a_i \geq 0,\ i \in \{1, \cdots, N\}$ are the Lagrange multipliers and $L(\mathbf{x}, \mathbf{a}) = F(\mathbf{x}) + \sum_{i=1}^{N} a_i f_i(\mathbf{x})$ is the Lagrangian.

# SVM: From Primal to Dual

Lagrange Relaxation

$$\min_{\mathbf{w},b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)}(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)) \right\}$$

Minimax Primal Formulation

Find $\mathbf{w}, b$ and $\mathbf{a}$ such that: Subject to: $a^{(n)} \geq 0, \ \ \forall(\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

$$\min_{\mathbf{w},b} \max_{\mathbf{a}} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)}(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)) \right\}$$

=

Dual Formulation

$$\max_{\mathbf{a}} \min_{\mathbf{w},b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)}(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)) \right\}$$

# SVM: From Primal to Dual

Lagrange Relaxation

$$\min_{\mathbf{w},b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)}\left(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)\right) \right\}$$

Minimax Primal Formulation

Find $\mathbf{w}, b$ and $\mathbf{a}$ such that: Subject to: $a^{(n)} \geq 0, \ \ \forall(\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

$$\min_{\mathbf{w},b} \max_{\mathbf{a}} \left\{ \underbrace{\frac{1}{2}\|\mathbf{w}\|^2}_{\text{Convex}} + \sum_{n=1}^{N} a^{(n)}\underbrace{\left(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)\right)}_{\text{Convex}} \right\}$$

=

Dual Formulation

$$\max_{\mathbf{a}} \min_{\mathbf{w},b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} a^{(n)}\left(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b)\right) \right\}$$

# Further Simplifying Equations

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

Once $\mathbf{a}$ is fixed, there are no constraints and, at the optimum, $\nabla L(\mathbf{w})$ equals to zero (KKT):

$$\mathbf{w} - \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi(\mathbf{x}^{(n)}) = 0 \longrightarrow \mathbf{w} = \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi(\mathbf{x}^{(n)})$$

And so does $\frac{\partial L}{\partial b}$: $\quad \sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$

# Further Simplifying Equations

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

Consider the following as a constraint to eliminate $b$:

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$$

Substituting

$$\mathbf{w} = \sum_{n=1}^{N} a^{(n)} y^{(n)} \phi(\mathbf{x}^{(n)})$$

See the book or lecture notes for a step-by-step on how to use the information above to eliminate $\mathbf{w}$ and $b$.

# The Dual Representation

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

Kernel function

$$\operatorname*{argmax}_{\mathbf{a}} \tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a^{(n)}a^{(m)}y^{(n)}y^{(m)}k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$$

Dual Formulation

Inner product

Where: $k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}) = \phi(\mathbf{x}^{(n)})^T\phi(\mathbf{x}^{(m)})$

Subject to: $a^{(n)} \geq 0, \forall n \in \{1, \cdots, N\}$ $\quad \sum_{n=1}^{N} a^{(n)}y^{(n)} = 0$

# Why Is This Useful?

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

$$\Downarrow$$

Kernel function

$$\underset{\mathbf{a}}{\operatorname{argmax}}\, \tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a^{(n)}a^{(m)}y^{(n)}y^{(m)}k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$$

Dual Formulation

Where: $k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}) = \phi(\mathbf{x}^{(n)})^T\phi(\mathbf{x}^{(m)})$

Subject to: $a^{(n)} \geq 0, \forall n \in \{1, \cdots, N\}$ $\quad \sum_{n=1}^{N} a^{(n)}y^{(n)} = 0$

# Why Is This Useful?

There is a way to compute $k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}) = \phi(\mathbf{x}^{(n)})^T \phi(\mathbf{x}^{(m)})$ without having to ever compute $\phi(\mathbf{x})$. This is called the Kernel Trick.

This will also give us further insights into the optimal hyperplane (discussed in the next lecture).

# Calculating $k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$: The Kernel Trick

$$\mathbf{x} = (x_1, x_2)^T \rightarrow \phi(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2, x_1 x_2)^T$$

$$\mathbf{x} = (x_1, x_2)^T \rightarrow \phi(\mathbf{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1 x_2)^T$$

$$k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}) = \phi(\mathbf{x}^{(n)})^T \phi(\mathbf{x}^{(m)})$$

$k(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x})^T \phi(\mathbf{z})$ Where $\mathbf{x} = \mathbf{x}^{(n)}$ and $\mathbf{z} = \mathbf{x}^{(m)}$.

$$= (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1 x_2)\ (1, \sqrt{2}z_1, \sqrt{2}z_2, z_1^2, z_2^2, \sqrt{2}z_1 z_2)^T$$

$$= 1 + 2x_1 z_1 + 2x_2 z_2 + x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1 x_2 z_1 z_2$$

$$= (1 + x_1 z_1 + x_2 z_2)^2$$

$$= (1 + \mathbf{x}^T \mathbf{z})^2$$

—> these are the original input variables!

# Creating Polynomial Kernels

- This calculation can be generalised to basis expansions composed of all terms of order up to $p$.

$$k(\mathbf{x}, \mathbf{z}) = (1 + \mathbf{x}^T \mathbf{z})^p$$

Primal Formulation

Depends on the dimension of the feature transform applied to each training example.
There are $d+1$ variables, where $d$ is the dimensionality of the embedding.

$$\underset{\mathbf{w},b}{\operatorname{argmin}} \left\{ \frac{1}{2} \boxed{\|\mathbf{w}\|^2} \right\}$$

Subject to: $y^{(n)} \left( \boxed{\mathbf{w}^T \phi(\mathbf{x}^{(n)})} + b \right) \geq 1$

$\forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T}$

Dual Representation

Depends on the number of training examples $N$.
There are $N$ variables.

$$\underset{\mathbf{a}}{\operatorname{argmax}} \, \tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2} \boxed{\sum_{n=1}^{N} \sum_{m=1}^{N}} a^{(n)} a^{(m)} y^{(n)} y^{(m)} k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$$

Where: $k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}) = \phi(\mathbf{x}^{(n)})^T \phi(\mathbf{x}^{(m)})$ e.g., $(1 + \mathbf{x}^T \mathbf{z})^p$

Subject to: $a^{(n)} \geq 0, \ \forall n \in \{1, \cdots, N\}$ $\quad \sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$

# Summary So Far…

- Based on the idea of Lagrange relaxation, we can go from the primal to the dual representation of the SVM optimisation problem.

- The dual representation can avoid having to compute the basis expansions (feature transformations) by using the kernel trick.

- This allows us to use very high dimensional embeddings.