COMPSCI 689
Lecture 10: Properties of SVC and More Optimization

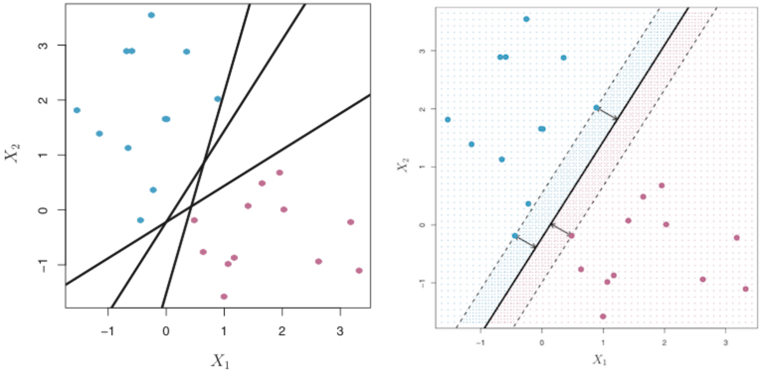
Benjamin M. Marlin

College of Information and Computer Sciences
University of Massachusetts Amherst

Slides by Benjamin M. Marlin (marlin@cs.umass.edu).

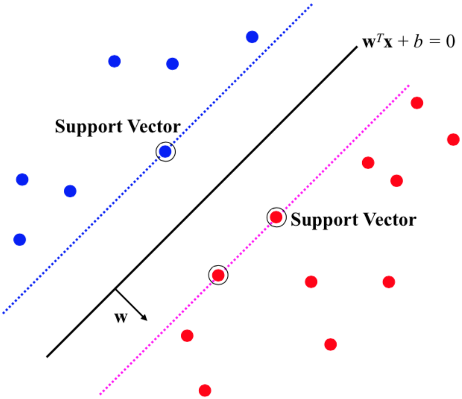
Maximum Margin Property

Part of popularity of SVMs stems from the fact that the hinge loss results in the *maximum margin* decision boundary when the training cases are linearly separable.



Support Vector Property

In the linearly separable case, some data points will always fall exactly on the margins. These points are called *support vectors* and they uniquely determine the optimal model parameters.



A General Constrained Problem

$$\begin{aligned} \mathbf{x}_* = \arg\min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} \quad & c_i(\mathbf{x}) = 0 \quad i \in \mathcal{E} \\ & c_i(\mathbf{x}) \geq 0 \quad i \in \mathcal{I} \end{aligned}$$

Our existing optimization tools only work for unconstrained problems. We need additional tools for dealing with constrained problems. Constrained convex problems have particularly interesting structure.

The Lagrangian

- The fundamental tool for analyzing constrained optimization problems is the Lagrangian function.
- Given an objective function $f(\mathbf{x})$ and a set of equality and inequality constraint functions $c_i(\mathbf{x}) = 0$ for $i \in \mathcal{E}$ and $c_i(\mathbf{x}) \geq 0$ for $i \in \mathcal{I}$, the Lagrangian is the function:

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i \in \mathcal{I}} \lambda_i c_i(\mathbf{x}) - \sum_{i \in \mathcal{E}} \lambda_i c_i(\mathbf{x})$$

- The new variables λ_i are referred to as *Lagrange Multipliers*.

Example: Lagrangian

Consider the constrained quadratic optimization problem shown below where $\mathbf{A}$ is assumed to be positive semi-definite.

$$\mathbf{x}_* = \arg\min_{\mathbf{x}} \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{b} + c$$
$$\text{s.t. } \mathbf{x}^T\mathbf{d} = f$$

What is the geometric interpretation of this optimization problem?

What is the Lagrangian for this optimization problem?

Ans: $\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{b} + c - \lambda(\mathbf{x}^T\mathbf{d} - f)$

Example: Lagrangian

Consider the constrained quadratic optimization problem shown below where $\mathbf{A}$ is assumed to be positive semi-definite.

$$\mathbf{x}_* = \arg\min_{\mathbf{x}} \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{b} + c$$
$$\text{s.t. } \mathbf{x}^T\mathbf{d} \geq f$$

What is the geometric interpretation of this optimization problem?

What is the Lagrangian for this optimization problem?

Ans: $\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{b} + c - \lambda(\mathbf{x}^T\mathbf{d} - f)$

The KKT Conditions

- The KKT conditions specify necessary first-order conditions on the solution $\mathbf{x}^*$ in terms of the Lagrangian $\mathcal{L}(\mathbf{x}, \lambda)$.
- Suppose that $\mathbf{x}_*$ is a solution to a constrained optimization problem with Lagrangian $\mathcal{L}(\mathbf{x}, \lambda)$ and constraints $c_i(\mathbf{x}) = 0$ for $i \in \mathcal{E}$ and $c_i(\mathbf{x}) \geq 0$ for $i \in \mathcal{I}$. Then there existis a λ_* such that:

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}_*, \lambda_*) &= 0 \\
c_i(\mathbf{x}_*) &= 0 \text{ for all } i \in \mathcal{E} \\
c_i(\mathbf{x}_*) &\geq 0 \text{ for all } i \in \mathcal{I} \\
\lambda_{i*} &\geq 0 \text{ for all } i \in \mathcal{I} \\
\lambda_{i*} c_i(\mathbf{x}_*) &= 0 \text{ for all } i \in \mathcal{I} \cup \mathcal{E}
\end{aligned}$$

Method of Lagrange Multipliers

- If we only have equality constraints, to find a point that satisfies the KKT conditions, we identify and analyze the stationary points of the Lagrangian by solving the Lagrangian gradient system:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = 0$$
$$\nabla_{\lambda} \mathcal{L}(\mathbf{x}, \lambda) = 0$$

- We identify all values of $\mathbf{x}$, and λ satisfying the above equations, plug the values of $\mathbf{x}$ into $f(\mathbf{x})$ and determine the minimizer.
- This is called the method of Lagrange multipliers.

Example: Method of Lagrange Multipliers

Consider the constrained problem shown below.

$$\mathbf{x}_* = \arg\min_{\mathbf{x}} x_1 + x_2$$
$$\text{s.t. } x_1^2 + x_2^2 = 2$$

What is the geometric interpretation of this optimization problem?

What is the Lagrangian for this optimization problem?

What is the solution to this optimization problem?

Example: Method of Lagrange Multipliers

The Lagrangian is $\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2)$. To find the optimizer, we solve the Lagrangian gradient system:

$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial x_1} = 1 - 2\lambda x_1 = 0 \Rightarrow x_1 = \frac{1}{2\lambda}$$
$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial x_2} = 1 - 2\lambda x_2 = 0 \Rightarrow x_2 = \frac{1}{2\lambda}$$
$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial \lambda} = x_1^2 + x_2^2 - 2 = 0 \Rightarrow x_1^2 + x_2^2 = 2$$

Plugging the first two results into the third result we find that $\lambda^2 = \sqrt{1/4}$ and thus $\lambda = \pm 1/2$. Plugging this back into the first two results we get that $x_1 = x_2 = \pm 1$. Checking both solutions by plugging into $x_1 + x_2$, the minimizer is $x_1 = x_2 = -1$.