Support Vector Machines

- Support vector machines can be viewed as the *dual* to maximum margin classifiers.
- Maximum margin classifiers represent optimization problems with the maximum margin between the supporting hyperplanes as the optimization criterion.
- A particularly convenient technique to derive the dual of an optimization problem is a technique referred as the *Lagrangian dual*.

Lagrangian Optimization

Assume that we have a *primal* optimization problem of the form,

$$\min_{\overline{x}} \phi(\overline{x}),$$

such that

$$g_i(\overline{x}) \geq 0,$$

for all $\overline{x} \in \mathbb{R}^n$ with $i = 1, \ldots, l$. Here we assume that ϕ is a convex objective function and we also assume that the constraints g_i are linear.

We can construct the *Lagrangian optimization problem* as follows,

$$\max_{\overline{\alpha}} \min_{\overline{x}} L(\overline{\alpha}, \overline{x}) = \max_{\overline{\alpha}} \min_{\overline{x}} \left(\phi(\overline{x}) - \sum_{i=1}^{l} \alpha_i g_i(\overline{x}) \right),$$

such that

$$\alpha_i \geq 0,$$

for $i = 1, \ldots, l$ and $\overline{x} \in \mathbb{R}^n$.

Lagrangian Optimization

Observations:

- The new objective function $L(\overline{\alpha}, \overline{x})$ is called the *Lagrangian* and incorporates the original objective function ϕ together with a linear combination of the constraints g_i.
- The values $\alpha_1, \ldots, \alpha_l$ are called the *Lagrangian multipliers*.
- We call $\overline{x}$ the *primal variable* and $\overline{\alpha}$ the *dual variable*.
- This newly derived optimization problem has the unusual feature of two nested optimization operators with opposing optimization objectives.

Lagrangian Optimization

We have

$$\max_{\overline{\alpha}} \min_{\overline{x}} L(\overline{\alpha}, \overline{x}) = \max_{\overline{\alpha}} \min_{\overline{x}} \left(\phi(\overline{x}) - \sum_{i=1}^{l} \alpha_i g_i(\overline{x}) \right),$$

now let $\overline{x} = \overline{x}^*$ be an optimum then

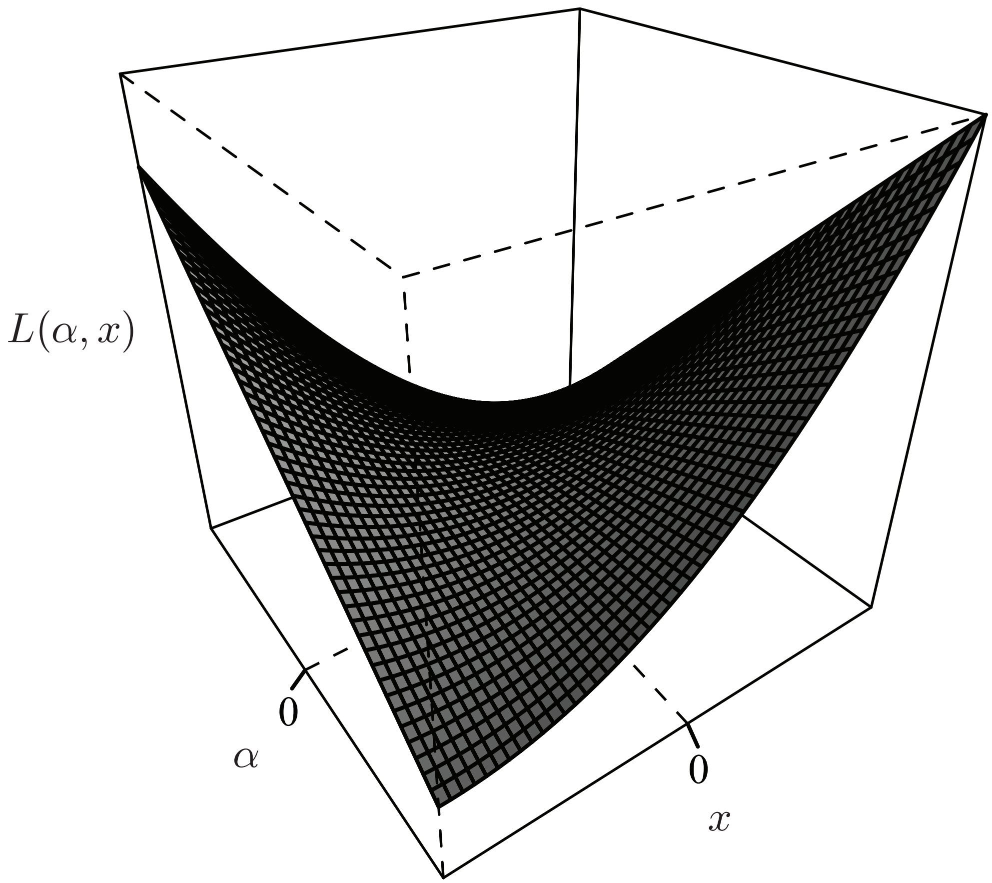
$$\max_{\overline{\alpha}} L(\overline{\alpha}, \overline{x}^*) = \max_{\overline{\alpha}} \left(\phi(\overline{x}^*) - \sum_{i=1}^{l} \alpha_i g_i(\overline{x}^*) \right),$$

now let $\overline{\alpha} = \overline{\alpha}^*$ be an optimum then

$$\min_{\overline{x}} L(\overline{\alpha}^*, \overline{x}) = \min_{\overline{x}} \left(\phi(\overline{x}) - \sum_{i=1}^{l} \alpha_i^* g_i(\overline{x}) \right).$$

This implies that our solutions are *saddle points* on the graph of the function $L(\overline{\alpha}, \overline{x})$.

Saddle Point



Lagrangian Optimization

An important observation is that at the saddle point the identity

$$\frac{\partial L}{\partial \overline{x}} = \overline{0},$$

has to hold.

This gives us the important identity

$$\frac{\partial L}{\partial \overline{x}}(\overline{\alpha}, \overline{x}^*) = \overline{0}.$$

Here, the point $\overline{x}^*$ represents an optimum of L with respect to $\overline{x}$.

Lagrangian Optimization

This allows us to formulate our first major result in Lagrangian optimization theory. Let $\overline{\alpha}^*$ and $\overline{x}^*$ be a solution to the Lagrangian such that,

$$\max_{\overline{\alpha}} \min_{\overline{x}} L(\overline{\alpha}, \overline{x}) = L(\overline{\alpha}^*, \overline{x}^*) = \phi(\overline{x}^*) - \sum_{i=1}^{l} \alpha_i^* g_i(\overline{x}^*),$$

then $\overline{x}^*$ is a solution to the primal objective function if and only if the following conditions hold,

$$\begin{aligned} \frac{\partial L}{\partial \overline{x}}(\overline{\alpha}^*, \overline{x}^*) &= \overline{0}, \\ \alpha_i^* g_i(\overline{x}^*) &= 0, \\ g_i(\overline{x}^*) &\geq 0, \\ \alpha_i^* &\geq 0, \end{aligned}$$

for $i = 1, \ldots, l$.

These conditions are collectively referred to as the *Karush-Kuhn-Tucker (KKT) conditions* and if satisfied ensure that

$$L(\overline{\alpha}^*, \overline{x}^*) = \phi(\overline{x}^*).$$ (Why?)

NOTE: The KKT conditions are always satisfied for convex optimization problems!

Lagrangian Dual

Now let $\overline{x}^*$ be an optimum, that is,

$$\frac{\partial L}{\partial \overline{x}}(\overline{\alpha}, \overline{x}^*) = \overline{0},$$

then we can rewrite our Lagrangian as an objective function of only the dual variable,

$$L(\overline{\alpha}, \overline{x}^*) = \phi'(\overline{\alpha}).$$

We call the function ϕ' the *Lagrangian dual*.

This gives us our new, dual optimization problem

$$\max_{\overline{\alpha}} \phi'(\overline{\alpha}),$$

subject to

$$\alpha_i \geq 0,$$

for $i = 1, \ldots, l$.

NOTE: Observe that

$$\max_{\overline{\alpha}} \phi'(\overline{\alpha}) = \phi'(\overline{\alpha}^*) = L(\overline{\alpha}^*, \overline{x}^*) = \phi(\overline{x}^*),$$

if the KKT conditions are satisfied.

An Example

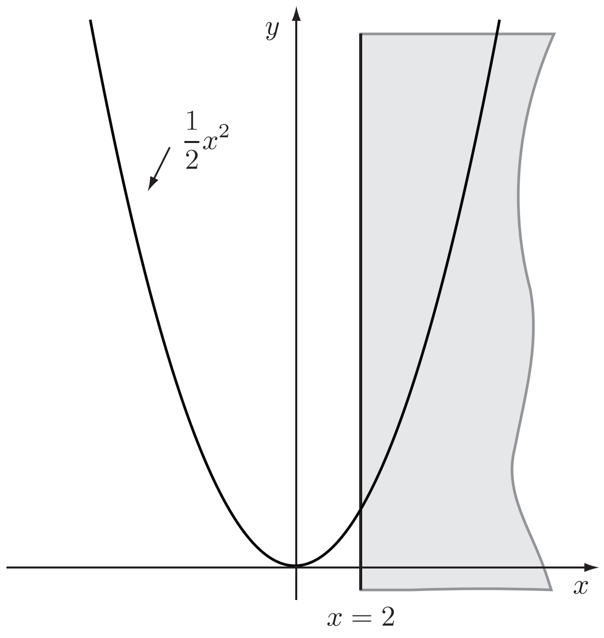
Consider the convex optimization problem,

$$\min \phi(x) = \min \frac{1}{2}x^2,$$

subject to the linear constraint

$$g(x) = x - 2 \geq 0,$$

with $x \in \mathbb{R}$.

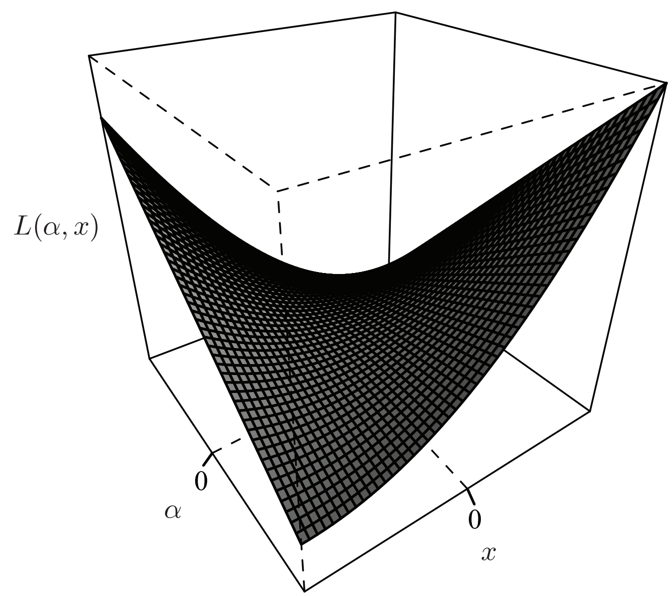


An Example

In order to solve this optimization problem using the Lagrangian dual we first construct the Lagrangian,

$$L(\alpha, x) = \frac{1}{2}x^2 - \alpha(x - 2).$$

As expected for a convex objective function we have a unique saddle point in the graph of the Lagrangian,



An Example

We know that this saddle point has to occur where the gradient of the Lagrangian with respect to the variable x is equal to zero,

$$\frac{\partial L}{\partial x}(\alpha, x^*) = x^* - \alpha = 0.$$

Solving for x^* gives us,

$$x^* = \alpha.$$

Now, plugging this identity back into the Lagrangian gives us,

$$L(\alpha, x^*) = \frac{1}{2}\alpha^2 - \alpha^2 + 2\alpha = 2\alpha - \frac{1}{2}\alpha^2.$$

This Lagrangian has no longer any dependencies on the variable x and therefore we can rewrite this as the Lagrangian dual with $\phi'(\alpha) = L(\alpha, x^*)$ or,

$$\max_{\alpha} \phi'(\alpha) = \max_{\alpha} \left(2\alpha - \frac{1}{2}\alpha^2\right)$$

subject to

$$\alpha \geq 0.$$

An Example

Now we know that $L(\alpha, x)$ is convex, therefore $\phi'(\alpha^*) = \max_\alpha \phi'(\alpha)$ implies that

$$\frac{d\phi'}{d\alpha}(\alpha^*) = 2 - \alpha^* = 0.$$

This means that,

$$x^* = \alpha^* = 2,$$

as required by our observation of the primal optimization problem.

Formally we can show that the solution to the primal optimization problem and to the Lagrangian dual must coincide by showing that that the second KKT condition is satisfied,

$$\alpha^* g(x^*) = \alpha^*(x^* - 2) = 2(2-2) = 0.$$