THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 385: Stat Models for Big Data

Lecture 5a: Duality

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

https://psarkar.github.io/teaching

Duality

- So far we were doing unconstrained optimization:

$$\min_x f_0(x)$$

- Often you will need to add constraints:

$$\min_x f_0(x) \qquad \text{s.t.} \quad f_i(x) \leq 0, i = 1, \ldots, m$$

Duality

- So far we were doing unconstrained optimization:

$$\min_x f_0(x)$$

- Often you will need to add constraints:

$$\min_x f_0(x) \qquad \text{s.t.} \quad f_i(x) \leq 0, i = 1, \dots, m$$

- Idea: turn this into an unconstrained optimization – how about optimizing the following instead:

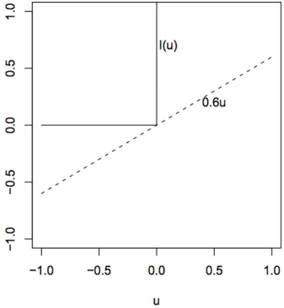
$$J(x) = \begin{cases} f_0(x) & \text{if } f_i(x) \leq 0, i = 1, \dots, m \\ \infty & \text{otherwise} \end{cases} = f_0(x) + \sum_i I(f_i(u))$$

Penalty

- $I(u)$ basically gives infinite penalty if $u > 0$

$$I(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

- Really messy formulation, non differentiable and discontinuous.



Proximal methods

- So, replacing the infinite penalty with a linear function gives the Lagrangian.

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$$

Proximal methods

- So, replacing the infinite penalty with a linear function gives the Lagrangian.

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$$

- So, if I maximize w.r.t $\lambda_i, \lambda_i \geq 0$ I get $J(x)$ – how?

Proximal methods

- So, replacing the infinite penalty with a linear function gives the Lagrangian.

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$$

- So, if I maximize w.r.t $\lambda_i, \lambda_i \geq 0$ I get $J(x)$ – how?
- If the constraints are satisfied, then set $\lambda_i = 0$
- If the i^{th} constraint is not satisfied, set $\lambda_i = \infty$

Proximal methods

- So, replacing the infinite penalty with a linear function gives the Lagrangian.

$$L(x, \lambda) = f_0(x) + \sum_i \lambda_i f_i(x)$$

- So, if I maximize w.r.t $\lambda_i, \lambda_i \geq 0$ I get $J(x)$ – how?
- If the constraints are satisfied, then set $\lambda_i = 0$
- If the i^{th} constraint is not satisfied, set $\lambda_i = \infty$
- Recall I wanted to minimize $J(x)$, so the problem becomes

$$\min_x \max_\lambda L(x, \lambda)$$

- Still tricky, but in many instances gets easier if we switch the order.

Dual

- So optimize

$$\max_{\lambda} \underbrace{\min_{x} L(x, \lambda)}_{g(\lambda)}$$

- $g(\lambda)$ is the dual function.
- the maximization over λ is known as the dual problem

Dual

- So optimize

$$\max_{\lambda} \underbrace{\min_{x} L(x,\lambda)}_{g(\lambda)}$$

- $g(\lambda)$ is the dual function.
- the maximization over λ is known as the dual problem
- Note that $g(\lambda)$ is concave, why?

Dual

- So optimize

$$\max_{\lambda} \underbrace{\min_{x} L(x, \lambda)}_{g(\lambda)}$$

- $g(\lambda)$ is the dual function.
- the maximization over λ is known as the dual problem
- Note that $g(\lambda)$ is concave, why?
- Since it is a point wise maximum over affine functions.
    - For a fixed x $L(x, \lambda)$ is essentially a linear function of the $\lambda's$

Duality gap

- Remember $\lambda u \leq I(u)$
- So $L(x, \lambda) \leq J(x)$ for $\lambda \geq 0$

Duality gap

- Remember $\lambda u \leq I(u)$
- So $L(x, \lambda) \leq J(x)$ for $\lambda \geq 0$
- So $g(\lambda) = \min_x L(x, \lambda) \leq min_x J(x) = p^*$ for $\lambda \geq 0$

Duality gap

- Remember $\lambda u \leq I(u)$
- So $L(x, \lambda) \leq J(x)$ for $\lambda \geq 0$
- So $g(\lambda) = \min_{x} L(x, \lambda) \leq min_X J(x) = p^*$ for $\lambda \geq 0$
- So $d^* = \max_{\lambda \geq 0} g(\lambda) \leq p^*$

Duality gap

- Remember $\lambda u \leq I(u)$
- So $L(x, \lambda) \leq J(x)$ for $\lambda \geq 0$
- So $g(\lambda) = \min_x L(x, \lambda) \leq min_x J(x) = p^*$ for $\lambda \geq 0$
- So $d^* = \max_{\lambda \geq 0} g(\lambda) \leq p^*$
- So, solving the dual is like finding the tightest lower bound on p^*

Duality gap

- Remember $\lambda u \leq I(u)$
- So $L(x, \lambda) \leq J(x)$ for $\lambda \geq 0$
- So $g(\lambda) = \min_x L(x, \lambda) \leq min_x J(x) = p^*$ for $\lambda \geq 0$
- So $d^* = \max_{\lambda \geq 0} g(\lambda) \leq p^*$
- So, solving the dual is like finding the tightest lower bound on p^*
- Strong duality: $d^* = p^*$
  - Holds if the optimization problem is convex, and a strictly feasible point exists, i.e. all constraints are satisfied and the inequality constraints are satisfied with strict inequalities.

Example – thanks to Vasko Lalkov and Jingya Li

$$\min(x_1^2 + x_2^2)$$
$$\text{s.t.} x_1 + x_2 \geq 4, x_1, x_2 \geq 0$$

Use the lagrangian:

$$\Lambda(x_1, x_2, \lambda, \nu_1, \nu_2) = x_1^2 + x_2^2 + \lambda(4 - x_1 - x_2) - \nu_1 x_1 - \nu_2 x_2$$

The dual is

$$g(\lambda, \nu_1, \nu_2) = \min_x \Lambda(x_1, x_2, \lambda, \nu_1, \nu_2) = 4\lambda + \min_{x_1}(x_1^2 - \lambda x_1 - \nu_1 x_1) + \min_{x_2}(x_2^2 - \lambda x_2 -$$

We get $2x_1^* = \lambda + \nu_1$ and $2x_2^* = \lambda + \nu_2$. So,

$$g(\lambda, \nu_1, \nu_2) = 4\lambda - (\lambda + \nu_1)^2/4 - (\lambda + \nu_2)^2/4$$

Cont.

Now we want:

$$\max_{\lambda \geq 0, \nu_1 \geq 0, \nu_2 \geq 0} g(\lambda, \nu_1, \nu_2)$$

Taking a derivative w.r.t ν_1, ν_2 and set it to zero.

$$\nu_1^* = \nu_2^* = -\lambda^* \qquad \Rightarrow \nu_1^* = \nu_2^* = 0$$

Taking a derivative w.r.t λ and set it to zero.

$$\lambda^* = (4 - \nu_1^*/2 - \nu_2^*/2) \Rightarrow \lambda^* = 4$$
$$\Rightarrow x_1^* = 2, x_2^* = 2$$

Reading

Look at the fantastic writeup by David Knowles on “Lagrangian Duality for Dummies”. I have linked this from the class website.