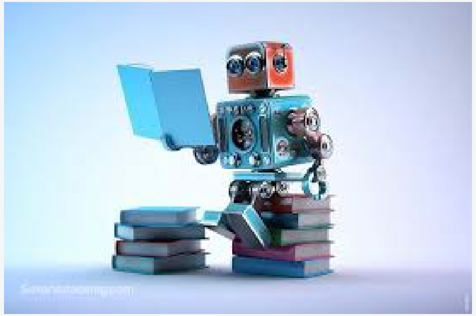
به نام خدا

دانشگاه علم و صنعت ایران

یادگیری ماشین



یادگیری ماشین

آرش عبدی هجراندوست

arash.abdi.hejrandoost@gmail.com

دانشگاه علم و صنعت

دانشکده مهندسی کامپیوتر

نیم سال اول ۱۴۰۱-۱۴۰۲

Dual Problem – مساله دوگان

- تکنیکی برای حل مسائل بهینه سازی که ممکن است راه حل ساده تری ارائه کند.

Primal Problem – مساله اصلی

$$p^* := \min_x \; f_0(x) \;\; : \;\; f_i(x) \leq 0, \;\; i = 1, \ldots, m.$$

$$f(x) := (f_1(x), \ldots, f_m(x)).$$

- مثلا یافتن حداقل فاصله تا یک چندوجهی:

$$p^* = \min_x \; \frac{1}{2}||x||_2^2 \;\; : \;\; Ax \leq b,$$

$$A \in \mathbf{R}^{m \times n}$$

$$b \in \mathbf{R}^m.$$

تابع لاگرانژ

$$L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$$

$$L(x, y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) = f_0(x) + y^T f(x).$$

$$y \in \mathbf{R}^m,$$

- y متغیر دوگان (Dual) یا متغیر/ضریب لاگرانژ است.
- مثال فاصله تا چندوجهی:

$$L(x, y) = \frac{1}{2}\|x\|_2^2 + y^T(Ax - b).$$

Dual Function – تابع دوگان

- مساله اصلی:

$$p^* := \min_x \; f_0(x) \;\; : \;\; f_i(x) \le 0, \;\; i = 1, \ldots, m.$$

- تابع لاگرانژ:

$$L(x, y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) = f_0(x) + y^T f(x).$$

- براساس تابع لاگرانژ، تابعی بر اساس صرفا متغیرهای دوگان (و نه متغیرهای اصلی) تعریف می‌کنیم که حدپایین از تابع هدف را ارائه می‌کند.
- برای مقدار ثابت متغیر دوگان $y \ge 0$، تابع لاگرانژ یک تابع هدف جریمه‌دار است:
  - جریمه خطی برای میزان تخطی از محدودیت های مساله اصلی
- جریمه صرفا وقتی مثبت است (اعمال جریمه صورت می‌گیرد) که حداقل یکی از محدودیت ها نقض شود.

تعریف تابع دوگان

$$L(x,y) = f_0(x) + \sum_{i=1}^{m} y_i f_i(x) = f_0(x) + y^T f(x).$$

✗ اگر همه محدودیت‌ها رعایت شوند، نقطه x نقطه‌ای مجاز (Feasible) است و داریم: $y^T f(x) \le 0$

$$\forall\, x \text{ feasible}, \;:\; f_0(x) \ge L(x,y).$$

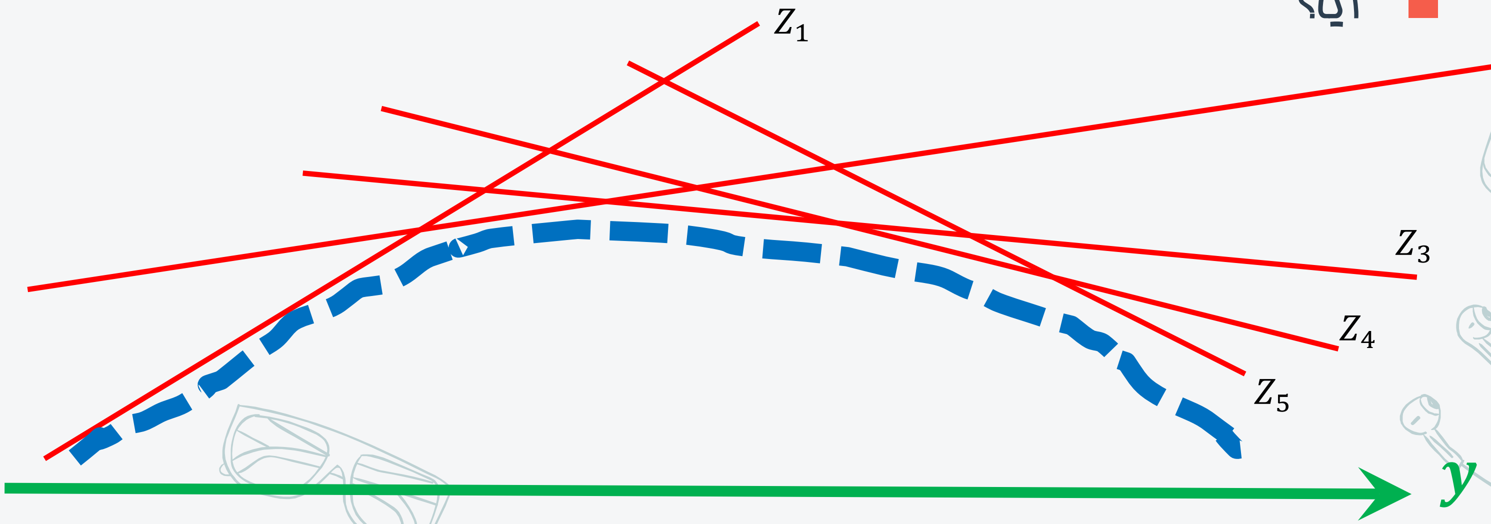
✗ تابع دوگان (dual function):

$$g(y) := \min_z \; L(z,y) = \min_z \; f_0(z) + \sum_{i=1}^{m} y_i f_i(z).$$

$$\forall x \text{ feasible}, \;:\; f_0(x) \ge L(x,y).$$

$$g(y) := \min_z \; L(z,y) = \min_z \; f_0(z) + \sum_{i=1}^{m} y_i f_i(z).$$

- تابع دوگان تابعی مقعر (Concave) است،
  - حتی اگر مساله اصلی محدب (Convex) نباشد
  - زیرا حاصل کمینه گیری نقطه به نقطه روی تعدادی تابع affine (خطی) است.
    - آیا؟



$$\forall x \text{ feasible}, \;:\; f_0(x) \geq L(x,y).$$

$$g(y) := \min_z L(z,y) = \min_z f_0(z) + \sum_{i=1}^{m} y_i f_i(z).$$

- برای هر x و y داریم: $L(x,y) \geq g(y)$
- و خواهیم داشت: $\forall\, x\ feasible,\ f_0(x) \geq g(y)$
- تابع دوگان g، حدپایینی از تابع بهینه سازی اصلی در مجموعه متغیرهای اصلی مجاز (feasible) ارائه می‌کند.
- $g(y)$ مستقل از x است.
- کمینه گیری از نامعادله بالا روی x خواهد داد:

$$\forall\, y \geq 0 : p^* \geq g(y)$$

محاسبه این حد پایین می‌تواند ساده باشد، خصوصا که محدودیتی در بهینه سازی ندارد.

مثالی از تابع دوگان

- مساله کمترین فاصله تا چند وجهی

$$g(y) = \min_x L(x, y) = \min_x \frac{1}{2}\|x\|_2^2 + y^T(Ax - b).$$

- مساله ای محدب است که با برابر صفر قراردادن مشتق جواب به دست می‌آید:

$$x^*(y) = -A^T y$$

- و با جایگذاری آن در تابع دوگان داریم:

$$g(y) = L(x^*(y), y) = -b^T y - \frac{1}{2}\|A^T y\|_2^2.$$

Dual Problem – مساله دوگان

- تابع دوگان g (برحسب متغیر دوگان y و **به ازای هر مقدار برای y**) حد پایینی برای مساله اصلی است.
- حدپایین مفیدتر، بزرگترین حد پایین است
- پس بهتر است روی y مقداری بیشینه را بیابیم:

$$p^* \geq d^* \;:=\; \max_{y \geq 0} g(y)$$

- مساله یافتن بیشینه در تابع دوگان (حدپایین)، مساله دوگان نام دارد.
- مقدار بهینه d^* مقدار بهینه دوگان است.
- از آنجا که g مقعر است، و محدودیت‌های روی y محدودیت‌های خطی هستند $(y \geq 0)$، بیشینه یابی فوق مساله ای محدب است.
  - ابزارهای مختلفی برای حل آن وجود دارد.

مثال

- در مثال قبل:

$$d^* = \max_{y \geq 0} g(y) = \max_{y \geq 0} -b^T y - \frac{1}{2}\|A^T y\|_2^2.$$

دوگان ضعیف یا قوی
Weak vs Strong Duality

- دوگان ضعیف:

$$p^* \geq d^*$$

- دوگان قوی:

$$p^* = d^*$$

- فاصله بین d^* و p^* را *Duality Gap* می‌نامند.
- در مسائل بهینه سازی *convex* تقریبا همیشه دوگان قوی برقرار است.

چند نکته:

- با این روش، حد پایین برای مسائل بهینه سازی به دست می‌آید
- نیازی نیست مساله اصلی convex باشد.
- ممکن است تابع g به سادگی قابل محاسبه نباشد.
  - مستلزم حل یک مساله بهینه سازی دیگر است.
- این روش وقتی مفیدتر است که تابع g به صورت فرم بسته (closed form) قابل محاسبه باشد.

استفاده در SVM

$$\underset{\mathbf{w},b}{\arg\min}\ \frac{1}{2}\|\mathbf{w}\|^2$$

$$t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right) \geqslant 1, \qquad n=1,\ldots,N.$$

- تعریف تابع لاگرانژ:

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

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

- برای تابع دوگان، لازم است برحسب متغیرهای اصلی w,b کمینه بگیریم
- بنابراین مشتق بر حسب w و b را برابر با صفر قرار می‌دهیم و داریم:

$$\mathbf{w} = \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n)$$

$$0 = \sum_{n=1}^{N} a_n t_n.$$

- با جایگذاری آنها، تابع دوگان بر حسب متغیرهای دوگان به دست می‌آید
  - که در آن استفاده از ترفند هسته نیز ممکن شده است.
- می‌خواهیم تابع دوگان زیر را بیشینه کنیم (مساله دوگان):

$$\widetilde{L}(\mathrm{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(\mathrm{x}_n, \mathrm{x}_m)$$

subject to

$$a_n \geqslant 0, \qquad n = 1, \ldots, N,$$

$$\sum_{n=1}^{N} a_n t_n = 0.$$

$$\mathbf{w} = \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n)$$

$$0 = \sum_{n=1}^{N} a_n t_n.$$

- برای دسته بندی (با جایگذاری w) داریم:

$$y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n k(\mathbf{x}, \mathbf{x}_n) + b.$$

- مساله دوگان فوق (بیشینه کردن تابع دوگان) شرایط Karush-Kuhn-Tucker (KKT) را طبق روابط زیر ارضا می‌کند:

$$\begin{aligned} a_n &\geqslant 0 \\ t_n y(\mathbf{x}_n) - 1 &\geqslant 0 \\ a_n \left\{ t_n y(\mathbf{x}_n) - 1 \right\} &= 0. \end{aligned}$$

- لذا طبق شرط سوم، برای هر داده:
  - یا ضریب لاگرانژ صفر است و لذا در محاسبه خروجی نقشی ایفا نمی‌کند
  - یا $t_n y(x_n) = 1$ که یعنی نقطه روی خط حداکثر حاشیه (margin) قرار دارد
    - SV است.

- برای محاسبه متغیر b، با توجه به اینکه برای SV ها داریم: $t_n y(x_n) = 1$
- میتوان معادله زیر را برای یکی از SV ها حل کرد:

$$t_n \left(\sum_{m \in S} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) + b \right) = 1$$

- برای پاسخ پایدارتر، با ضرب طرفین در t_n و توجه به اینکه $t_n^2 = 1$، و با متوسط گیری روی تمام SV ها داریم:

$$b = \frac{1}{N_S} \sum_{n \in S} \left(t_n - \sum_{m \in S} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) \right)$$