یادگیری ماشین

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

مدرس: دکتر سید ابوالفضل مطهری

دانشکده‌ی مهندسی کامپیوتر

تمرین دوم

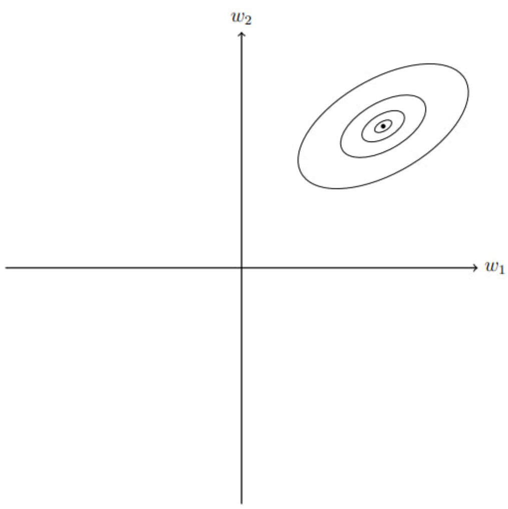
مسئله‌ی ۱. L-∞ Regularizer

نرم بینهایت بردار x را به این صورت $L_\infty(x) = max(|x_i| : i = \{۱, ۲, ..., n\})$ تعریف کنیم.

برای مثال اگر $x = [-۶, ۲, ۴]$ باشد، $L_\infty(x) = ۶$. همانطور که می‌دانید در مسائل یادگیری ماشین معمولا یک جمله regularization را به تابع هدف اضافه می‌کنیم.

الف

داخل نمودار loss پایین نرم L_∞ را بکشید و نقطه‌ای را مشخص کنید که بین مجموعه‌ای از وزن‌ها توسط این جمله regularization انتخاب می‌شود.



ب

استفاده از این نرم در فرایند یادگیری چه تاثیری روی وزن‌ها می‌گذارد؟ در چه شرایطی استفاده از این روش انتخاب مناسبی است؟

مسئله‌ی ۲. Regularization

فرض کنید برای بدست‌آوردن ضرایب در یک مسئله رگرسیون خطی، عبارت زیر را کمینه می‌کنیم:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \qquad \text{s.t} \qquad \sum_{j=1}^{p}|\beta_j| \leqslant s.$$

برای قسمت‌های زیر مشخص کنید کدام یک از حالت های ۱ تا ۵ رخ می‌دهد:

۱. اول افزایش و سپس کاهش می‌یابد.

۲. اول کاهش و سپس افزایش می‌یابد.

۳. افزایش می‌یابد.

۴. کاهش می‌یابد.

۵. ثابت می‌ماند.

الف

RSS(Residual Sum of Squares)، آموزش وقتی مقدار s را از صفر افزایش میدهیم.

ب

واریانس در حالت الف.

ج

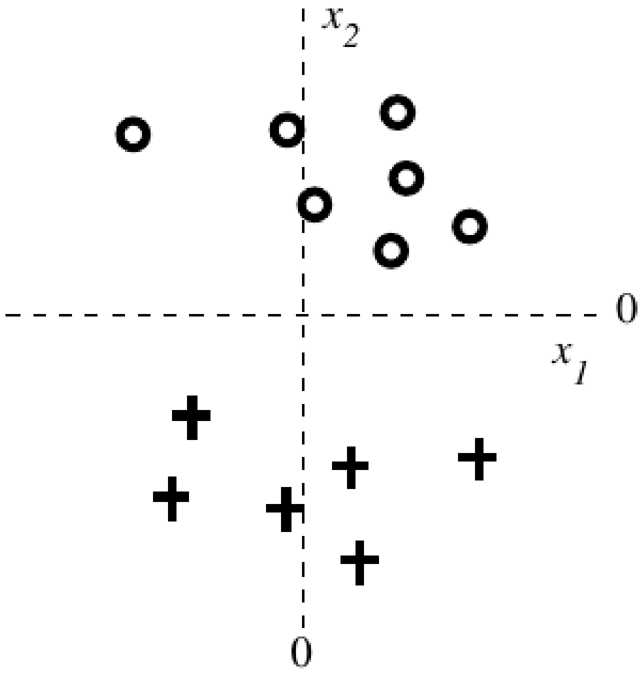
مربع بایاس در حالت الف.

مسئله‌ی ۳. Data Augmentation

- تاثیر روش های کاهش ابعاد مانند PCA یا هر روش دیگری را که میشناسید، بر overfitting بررسی کنید.
- می دانیم یک روش مناسب برای مقابله با overfitting اضافه کردن داده‌های جدید است؛ البته همیشه این کار امکان پذیر نیست. ولی می‌توانیم با استفاده از تکنیک های data augmentaion مجموعه داده خود را بزرگ‌تر کنیم و با overfitting مقابله کنیم. در مورد روش های data augmentation مطالعه کنید و چکیده یافته‌هایتان را بنویسید.
(۵ نمره امتیازی در صورتی که بتوانید برای این سوال تحلیل تئوری نیز انجام دهید).

مسئله‌ی ۴. Logistic Regression

مسئله دسته‌بندی دوتایی [1] در تصویر زیر را در نظر بگیرید.



برای این مسئله از یک مدل رگرسیون لجستیک ساده به صورت زیر استفاده کرده ایم.

$$P(y=۱|\vec{x},\vec{w}) = g(w_۰ + w_۱x_۱ + w_۲x_۲) = \frac{۱}{۱+\exp(-w_۰ - w_۱x_۱ - w_۲x_۲)}$$

دقت کنید که میتوان داده ها را با یک خط با خطای تمرین ۰ از هم جدا کرد.

فرض کنیم میخواهیم برای حل مسئله معادله regularized زیر را بیشینه کنیم.

$$\sum_{i=۱}^{n} \log(\mathbb{P}(y_i|x_i, w_۰, w_۱, w_۲)) - Cw_j^۲$$

به طوری که داریم: $j \in \{۰,۱,۲\}$.

به ازای مقادیر بزرگ C مشخص کنید که با انتخاب j های مختلف، خطای تمرین چه تغییری می‌کند؟

[1] binary classification

مسئله‌ی ۵. Generalized Linear Models (امتیازی ۱۰ نمره)

در این مسئله به دنبال یک فرم کلی تر برای خانواده توزیع نمایی می‌گردیم. ابتدا توزیع نرمال را در نظر داشته باشید:

$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{1}{2\sigma^2}(y-\mu)^2\}$$

.

الف

نشان دهید توزیع نرمال (بدون در نظر گرفتن واریانس برابر با ۱) خود یک توزیع نمایی است.به طور خاص موارد $b(y)$ ، η ، $T(y)$ ، $a(\eta)$ را در تابع توزیع احتمال نمایی مشخص کنید. رابطه توزیع احتمال نمایی به صورت زیر است:

$$p(y;\eta) = b(y) exp\{\eta^\top T(y) - a(\eta)\}$$

.

راهنمایی: از آنجا که σ^2 در اینجا یک متغیر است، η و $T(y)$ بردار های دو بعدی خواهند بود. برای ثابت‌بودن نمادگذاری فرض کنید: $\eta = [\eta_1 \ \ \eta_2]^\top$. همچنین $a(\eta)$ را برحسب η_1 و η_2 بنویسید.

ب

فرض کنید یک مجموعه آموزشی [2] با توزیع .i.i.d به صورت روبرو $\{(x^{(i)}, y^{(i)}), i = 1, ..., m\}$ داریم. با شروع از فرم کلی تابع توزیع احتمال نمایی ای که در بخش قبل بیان شد، فرم کلی ماتریس هِسیَن log-likelihood $l(0) = \sum_{i-1}^{m} \log \ p(y^{(i)}|x^{(i)}; 0)$ را بدست آورید. جواب باید برحسب x ، η_1 و η_2 باشد.

[2] training set