



باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

مقدمه ای بر یادگیری ماشین - گروه دکتر امینی

پاییز ۱۴۰۱

تمرین سری دوم

۱. مهلت تحویل این تمرین مطابق تاریخ اعلام شده در سامانه CW می باشد.

۲. ۱۴ روز تاخیر مجاز برای تحویل تمارین در اختیار شما خواهد بود.

۳. سقف تاخیر برای تحویل هر تمرین ۷ روز خواهد بود و پس از آن پاسخنامه تمرین منتشر خواهد شد.

۴. ابهامات و مشکلات خود درباره سه سوال اول تئوری را می‌توانید با آقای شمشیری و برای سه سوال دیگر با خانم سوداگر مطرح کنید.

@PouyaSha

@zahrasodagar

۱. مثالی از مدل گوسی

مدل زیر را در نظر بگیرید:

$$Y_i = 10.0 + 0.5X_i + \epsilon_i$$

$$\epsilon_i \sim_{(iid)} N(0, 1)$$

۱. برای این مدل مقادیر $E[Y|X=0]$ ، $E[Y|X=1]$ و $var[Y|X]$ را بدست آورید.

۲. اگر ورودی این مدل $X = 2$ باشد، چقدر احتمال دارد که خروجی آن مقداری بزرگتر از ۱۰ باشد؟

۳. فرض کنید ورودی‌های این مدل توزیعی گوسی با میانگین صفر و واریانس ۲۰ داشته باشند، مقدار میانگین و واریانس خروجی و کوواریانس متغیر تصادفی ورودی و خروجی را محاسبه کنید.

۲. رگرسیون مناسب

برای هر یک از موارد زیر، استدلال کنید که کدام یک از گزاره‌های الف، ب، ج یا د صحیح می‌باشد.

۱. مسئله Lasso نسبت به مسئله کوچکترین مربعات ...

۲. مسئله رگرسیون Ridge نسبت به مسئله کوچکترین مربعات ...

۳. روش‌های غیر خطی نسبت به مسئله کوچکترین مربعات ...

(الف) انعطاف‌پذیرتر است، بنابراین اگر مقدار افزایش بایاس کمتر از کاهش در واریانس باشد، دقت افزایش می‌یابد.

(ب) انعطاف‌پذیرتر است، بنابراین اگر افزایش در واریانس کمتر از کاهش بایاس باشد، دقت افزایش می‌یابد.

(ج) کمتر انعطاف‌پذیر است، بنابراین اگر افزایش در بایاس کمتر از کاهش واریانس باشد، دقت افزایش می‌یابد.

(د) کمتر انعطاف‌پذیر است، بنابراین اگر افزایش در واریانس کمتر از کاهش بایاس باشد، دقت افزایش می‌یابد.

۳ رگرسیون و تکیین بودن پاسخ

یکی از روش‌های رگولاریزه کردن در مسائل رگرسیون خطی روش Lasso می‌باشد. در این روش مقدار نرم L1 متغیر مسئله با ضریبی به تابع هزینه اضافه می‌شود. استفاده از این جمله جریمه باعث می‌شود پاسخ مسئله بهینه‌سازی بدست آمده تکیین‌تر باشد و بتوان از مقادیر کوچک‌تر و کم‌اهمیت در بردار بدست آمده بدون ایجاد خطایی زیاد صرف نظر کرد. در این قسمت از تمرین با هم به بررسی این می‌پردازیم که این جمله چگونه باعث تکیین‌تر شدن پاسخ مسئله می‌شود.

در ماتریس $X \in \mathbf{R}^{n \times d}$ هر ردیف مربوط به یک داده یادگیری می‌باشد (که هرکدام در مجموع d ویژگی دارند) و بردار $y \in \mathbf{R}^n$ مربوط به خروجی‌های مورد نظر و $w \in \mathbf{R}^d$ پارامتر وزن رگرسیون بوده و که w^* مقدار بهینه آن می‌باشد. برای سادگی محاسبات فرض کنید داده‌های آموزش سفید شده‌اند به طوری که رابطه $X^T X = I$ برقرار باشد. در رگرسیون Lasso بردار بهینه از رابطه زیر بدست می‌آید:

$$w^* = \operatorname{argmin}_w J_\lambda(w),$$

$$J_\lambda(w) = \frac{1}{2} \|y - Xw\|_2^2 + \lambda \|w\|_1 \quad (\lambda > 0)$$

۱. ابتدا نشان می‌دهیم که سفید کردن داده آموزش باعث مستقل شدن ویژگی‌های داده‌های آموزش می‌شود به طوری که می‌توان به طور مستقل w_i^* را تنها از ویژگی i ام و خروجی تعیین کرد. برای نشان دادن این موضوع، ابتدا نشان دهید که تابع $J_\lambda(w)$ را می‌توان به فرم

$$J_\lambda(w) = g(y) + \sum_{i=1}^d f(X_{.i}, y, w_i, \lambda),$$

نوشت، به طوری که $X_{.i}$ ستون i ام ماتریس ویژگی بوده و تابع g تنها تابع متغیر y بوده و تابع f تابع $X_{.i}, y, w_i, \lambda$ می‌باشد.

۲. اگر $w_i^* > 0$ باشد، مقدار w_i^* را بدست آورید.

۳. اگر $w_i^* < 0$ باشد، مقدار w_i^* را بدست آورید.

۴. با توجه به دو قسمت قبل، در چه شرایطی مقدار w_i^* برابر با صفر می‌شود؟ چگونه می‌توان این شرط را اعمال کرد؟

۵. می‌دانیم که در مسئله رگرسیون Ridge، جمله جریمه در تابع هزینه به شکل $\frac{1}{2} \lambda \|w\|_2^2$ می‌باشد. در این حالت در چه شرایطی به پاسخ $w_i^* = 0$ می‌رسیم؟ تفاوت این شرط با قسمت قبل چه می‌باشد؟

۴ آنالیز بیزی توزیع نمایی

طول عمر یک ماشین را می‌توان به شکل متغیر تصادفی X از توزیع نمایی با پارامتر مجهول θ مدل کرد، به طوری که $p(x | \theta) = \theta e^{-\theta x}$ برای $x \geq 0$ و $\theta > 0$ برقرار باشد.

۱. نشان دهید که MLE برابر با $\hat{\theta} = 1/\bar{x}$ به طوری که $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ می‌باشد.

۲. فرض کنید که مقادیر $X_1 = 5$ ، $X_2 = 6$ و $X_3 = 4$ (مقادیر طول عمر سه ماشین از یک توزیع که از هم مستقل می‌باشند به واحد سال) مشاهده شده‌اند. مقدار MLE چقدر بدست می‌آید؟

۳. فرض کنید که می‌دانیم پارامتر θ نیز از توزیع نمایی به شکل $p(\theta) = \operatorname{Expon}(\theta | \lambda)$ بدست می‌آید. پارامتر اولیه $\hat{\lambda}$ را بگونه‌ای انتخاب کنید به طوری که $\mathbb{E}[\theta] = 1/3$ شود. (راهنمایی: توزیع گاما به فرم $\operatorname{Ga}(\theta | a, b) \propto \theta^{a-1} e^{-\theta b}$ می‌باشد به طوری که امید ریاضی آن a/b می‌باشد).

۴. توزیع پسین $p(\theta | \mathcal{D}, \hat{\lambda})$ را بدست بیاورید.

۵. امید ریاضی توزیع پسین $\mathbb{E}[\theta | \mathcal{D}, \hat{\lambda}]$ را بدست بیاورید.

۶. توضیح دهید که چرا MLE و امید ریاضی توزیع پسین متفاوت می‌باشند و در این مثال، استفاده از کدام یک برای تخمین پارامتر توزیع منطقی‌تر می‌باشد.

۵ Naïve Bayes با ویژگی‌های مخلوط

یک طبقه‌بندی Naïve Bayes سه کلاسه که دو ویژگی دارد را در نظر بگیرید، به طوری که یکی از این ویژگی‌ها باینری بوده و از توزیع برنولی می‌آید و ویژگی دیگر توزیعی گوسی دارد به طوری که:

$$y \sim \text{Mu}(y | \pi, 1), \quad x_1 | y = c \sim \text{Ber}(x_1 | \theta_c), \quad x_2 | y = c \sim \mathcal{N}(x_2 | \mu_c, \sigma_c^2)$$

$$\pi = (0.5, 0.25, 0.25), \quad \theta = (0.5, 0.5, 0.5), \quad \mu = (-1, 0, 1), \quad \sigma^2 = (1, 1, 1)$$

۱. $p(y | x_1 = 0, x_2 = 0)$ را محاسبه کنید (پاسخ باید برداری سه‌تایی باشد که مجموع سه درآیه آن برابر با ۱ شود).

۲. $p(y | x_1 = 0)$ را محاسبه کنید.

۳. $p(y | x_2 = 0)$ را محاسبه کنید.

۴. الگوی جالبی که در جواب‌های خود مشاهده می‌کنید را توجیه کنید (راهنمایی: به بردار پارامترهای θ دقت کنید).

۶ مرزهای تصمیم‌گیری گوسی

فرض کنید $p(x | y = j) = \mathcal{N}(x | \mu_j, \sigma_j^2)$ باشد به طوری که $j = 1, 2$ و $\mu_2 = 1, \sigma_1^2 = 1, \mu_1 = 0$ و $\sigma_2^2 = 10^6$ و احتمال هر کلاس $p(y = 1) = p(y = 2) = 0.5$ باشد.

۱. ناحیه تصمیم‌گیری $R_1 = x : p(x | \mu_1, \sigma_1) \geq p(x | \mu_2, \sigma_2)$ را بدست آورده و طرحی از این ناحیه‌ها رسم کنید (راهنمایی: هردو پاسخ معادله $p(x | \mu_1, \sigma_1) = p(x | \mu_2, \sigma_2)$ را بدست بیاورید).

۲. حال فرض کنید که $\sigma_2^2 = 1$ باشد. ناحیه R_1 را در این حالت بدست بیاورید.

۷ تمرین‌های کامپیوتری

در پوشه تمرین سه سوال CHW2_Q1، CHW2_Q2 و CHW2_Q3 قرار گرفته است. نوت‌بوک مربوط به هر سوال را تکمیل کرده و در فایل زیپ ارسالی نهایی خود قرار دهید.