

بنام او که عزت از آن اوست

تمرین شماره یک

دانشجویان گرامی:

این تمرین شامل مباحث Regression و classification میشود.

۱- مفهومی: درستی گزاره های زیر را بررسی کنید (با توضیح مختصر)

- (الف) خطای اعتبارسنجی (Validation) همواره کمتر از خطای آزمون است.
- (ب) افزایش پیچیدگی مدل در رگرسیون همیشه سبب افزایش واریانس و کاهش بایاس می شود.
- (پ) افزایش پیچیدگی مدل در رگرسیون همیشه سبب کاهش خطای آموزش و افزایش خطای آزمون می شود.
- (ت) در صورتی که مقدار بایاس بالا باشد و واریانس کم باشد، باید پیچیدگی مدل را کاهش داد.
- (ث) کاهش خطای آموزش منجر به کاهش خطای آزمون می شود.
- (ج) در صورتی که واریانس بالا باشد، افزایش داده های آموزش همیشه می تواند سبب کاهش واریانس شود.
- (چ) در صورتی که بایاس بالا باشد، افزایش داده های آموزش هیچ کمکی به کاهش بایاس نمی کند و باید پیچیدگی مدل را افزایش داد.

۲- رگرسیون خطی

(الف) فرض کنید داده های اندازه گیری شده برای سیستم خطی $y = 2x - 1$ که در آن مقدار نویز جمع شونده هم وجود دارد به صورت زیر باشد:

$$\{(-1,1), (0,8), (1,5), (-2,-1)\}$$

۱- نقاط داده شده و تابع مطلوب را ترسیم کنید.

۲- با استفاده از تابع هزینه SSE، خط برازش کننده را به دست آورید و آن را رسم کنید.

(ب) اگر w^* خط بهینه طبق معیار MSE (روی کل توزیع) را مشخص کند:

$$w^* = \operatorname{argmin}_w E_{x,y}[(y - w^T x)^2]$$

۱- w^* را بر حسب ماتریس خود همبستگی $R = E_x[xx^T]$ و بردار همبستگی $c = E_{xy}[xy]$ محاسبه نمایید.

۲- نشان دهید:

$$E_{x,y}[(y - w^{*T} x)x] = 0$$

۳- با استفاده از رابطه ی به دست آمده در بند قبل نشان دهید:

$$E_{x,y}[(y - w^{*T} x)(\hat{w}^T x - w^{*T} x)] = 0$$

۴- تعبیر هندسی رابطه ی بند ۳ چیست؟

۳- پیاده سازی

برای انجام این تمرین می بایست ابتدا کدی برای تولید داده به مقصد فایل **data** بنویسید. (این کد نیز جزء تمرین تحویلی است) سپس در ادامه از داده های این فایل استفاده نمایید، بطوریکه بتوان کد ها را روی یک فایل داده جایگزین اجرا کرد. در این فایل ۱۰۰ نمونه و مقادیر متناظر با آن ها به ترتیب در قالب دو ماتریس X و T قرار دهید. این داده ها را توسط رابطه ی $t = x^3 - 3x^2 - x + 3 + \epsilon$ تولید کنید که ϵ نویز گوسی جمع شونده با میانگین صفر و واریانس ۲ باشد.

برای بررسی الگوریتم ها، از بین این داده ها n نمونه به صورت تصادفی به عنوان مجموعه ی آموزشی انتخاب شده و بقیه برای تست مورد استفاده قرار خواهند گرفت. (مقدار n در هر قسمت بیان خواهد شد). برای محاسبه ی خطا از میانگین $RMSE$ در ۱۰۰ بار اجرای الگوریتم استفاده کنید (که باید در هر بار اجرا n نمونه ی آموزش دوباره به صورت تصادفی انتخاب شوند). برای حل سوال های زیر سه تابع تبدیل زیر را در نظر بگیرید (در هر قسمت بیان شده است که از کدام توابع باید استفاده کنید):

$$\begin{aligned}\phi_1(x) &= [1 \ x \ x^2 \ x^3] \\ \phi_2(x) &= [1 \ x \ x^2 \dots x^{10}] \\ (\phi_3(x))_i &= \exp\left(-\frac{x - c_i}{2\sigma^2}\right); \quad c_i \in \{-3, -2.9, -2.8, \dots, 2.9, 3\}, \sigma = 1\end{aligned}$$

در این قسمت به پیاده سازی و بررسی انواع روش های رگرسیون خواهید پرداخت. برای این منظور از داده های موجود در فایل **data** استفاده خواهید کرد.

برای بررسی الگوریتم ها، از بین این داده ها n نمونه به صورت تصادفی به عنوان مجموعه ی آموزشی انتخاب شده و بقیه برای تست مورد استفاده قرار خواهند گرفت. (مقدار n در هر قسمت بیان خواهد شد). برای محاسبه ی خطا از میانگین $RMSE$ در ۱۰۰ بار اجرای الگوریتم استفاده کنید (که باید در هر بار اجرا n نمونه ی آموزش دوباره به صورت تصادفی انتخاب شوند). برای حل سوال های زیر سه تابع تبدیل زیر را در نظر بگیرید (در هر قسمت بیان شده است که از کدام توابع باید استفاده کنید):

$$\begin{aligned}\phi_1(x) &= [1 \ x \ x^2 \ x^3] \\ \phi_2(x) &= [1 \ x \ x^2 \dots x^{10}] \\ (\phi_3(x))_i &= \exp\left(-\frac{x - c_i}{2\sigma^2}\right); \quad c_i \in \{-3, -2.9, -2.8, \dots, 2.9, 3\}, \sigma = 1\end{aligned}$$

توضیح: طبق تعریف اخیر، تابع ϕ_3 یک تابع با خروجی ۶۱ بعدی را نشان می دهد که هر بعد آن متناظر با یک تابع گوسی است که مراکز این گوسی ها به صورت یکنواخت در بازه ی $[-3, 3]$ پراکنده شده و واریانس هایشان یکسان است.

الف) در صورتی که برای رگرسیون از تابع هزینه SSE

$$J(w) = \sum_{i=1}^n \left(y^{(i)} - w^T \phi(x) \right)^2 + \lambda \|w\|_2^2$$

و تابع تبدیل ϕ_2 را در نظر بگیرید،

۱- به ازای $n = 30$ و $\lambda \in \{10^{-8}, 10^{-7}, \dots, 10^2, 10^3, 10^4\}$ بردار w را محاسبه کرده، نمودار خطای آموزش و آزمون را به ازای مقادیر مختلف λ رسم نمایید. (در رسم نمودار، از $\ln \lambda$ به عنوان محور افقی استفاده نمایید.)

۲- جهت تعیین پارامتر λ در محدوده مقادیری که در قسمت قبل ذکر شد، روش k -fold cross validation را پیاده سازی کنید. سپس با استفاده از 10-fold CV نمودار میانگین خطا روی داده های آموزش، اعتبارسنجی و آزمون را به ازای مقادیر مختلف λ رسم نمایید. سپس بهترین مقدار λ را طبق این نمودار پیدا کنید. همچنین مقدار میانگین خطا روی داده های اعتبارسنجی را برای بهترین λ گزارش نمایید.

ب) در این بند، هدف بررسی بایاس و واریانس است. برای این منظور لازم است $L = 50$ مجموعه ی آموزشی D^1, \dots, D^L تشکیل شود که برای تشکیل هر کدام از این مجموعه ها باید n داده ی آموزش مطابق رابطه ی $t = x^3 - 3x^2 - x + 3 + \epsilon$ ایجاد شود که توزیع x یکنواخت در بازه $[-3, 3]$ و $\epsilon \sim N(0; 2)$ باشد.

۱- با فرض استفاده از تابع هزینه SSE با جمله ی منظم سازی $\|w\|_2^2$ و به کارگیری تابع تبدیل ϕ_2 ، برای هر کدام از مجموعه های آموزش D^1, \dots, D^L با اندازه $n = 20$ و $\lambda \in \{10^{-8}, 10^{-7}, \dots, 10^2, 10^3, 10^4\}$ پارامتر w را به دست آورید. مقدار $\left(\text{بایاس} \right)^2$ ، واریانس و میانگین خطای آزمون را به ازای مقادیر مختلف λ روی یک نمودار رسم کنید. (راهنمایی: می توانید برای محاسبه ی بایاس و واریانس به روابط $3,47 - 3,45$ کتاب Bishop مراجعه نمایید.)

۲- مشابه بند ۱ این بار در حالتی که اندازه مجموعه آموزش $n = 200$ باشد، نتایج را مجدداً به دست آورده، نمودارهای مربوطه را رسم کنید. سپس نتایج حاصل را با بند ۱ مقایسه و تحلیل نمایید.

-4

دسته بندی خطی

الف) ابتدا نشان دهید تابع هزینه پرسپترون را می توان به صورت $\sum_{i=1}^n \max(0, -y^{(i)} w^T x^{(i)})$ نوشت و سپس تابع ضرر (loss) متناظر با آن را در حالت $y = 1$ بر حسب $w^T x$ رسم نمایید. نتیجه را با تابع ضرر logistic regression (در حالت $y = 1$ بر حسب $w^T x$) مقایسه کنید و تأثیر این تفاوت را در عملکرد روش های مربوطه مشخص نمایید.

ب) در مسأله ی دسته بندی چند دسته ای، یک مجموعه نمونه "خطی جدایی پذیر" (linearly separable) گفته می شود اگر یک ماشین خطی (linear machine) وجود داشته باشد که همه ی نمونه های این مجموعه را درست دسته بندی کند. در حالتی که نمونه های هر دسته قابل جدا شدن از نمونه های بقیه ی دسته ها توسط یک ابر صفحه باشند، نمونه ها "کاملاً خطی جدایی پذیر" (totally linearly separable) گفته می شوند. همچنین این مجموعه نمونه "دو به دو خطی جدایی پذیر" گفته می شود اگر نمونه های هر زوج دسته توسط یک ابر صفحه قابل جدا شدن باشند.

۱- نشان دهید نمونه های کاملاً جدایی پذیر خطی، جدا پذیر خطی نیز هستند، اما عکس آن لزوماً درست نیست.

۲- نشان دهید نمونه های دو به دو خطی جدایی پذیر، لزوماً خطی جدایی پذیر نیستند. آیا بالعکس نمونه های خطی جدایی پذیر لزوماً دو به دو خطی جدایی پذیر نیز هستند؟

پ) فرض کنید به دنبال پیدا کردن راستایی هستیم که جدا سازی بین نمونه های دو دسته (وقتی که روی آن راستا افکنده می شوند) بیشینه شود و برای این منظور بیشینه کردن فاصله بین میانگین دو دسته را (بدون در نظر گرفتن پراکندگی داده ها) مد نظر قرار داده باشیم. در حالتی که قید $w^T w = 1$ هم لحاظ شود، نشان دهید جواب مسأله ی بهینه سازی بالا به صورت زیر در می آید:

$$w \propto (\mu_2 - \mu_1)$$

همچنین توضیح دهید قید $w^T w = 1$ به چه منظور لحاظ شده است.

پیروز باشید