دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

پاییز و زمستان ۱۴۰۴
استاد: علی شریفی زارچی

گردآورندگان:

حل‌تمرین ۱ | یادگیری با نظارت | مهلت ارسال:

۱. (نمره) مسئله‌ای داریم که در آن داده‌های واقعی از تابع $f(x)$ تولید می‌شوند. با استفاده از مجموعه‌ی داده‌های یادگیری $\mathcal{D}$، که مقادیر آن از همان تابع $f(x)$ پیروی می‌کنند، مدل $g^{(\mathcal{D})}(x)$ را آموزش می‌دهیم. در این صورت خطای مورد انتظار برای داده‌های خارج از نمونه از رابطه‌ی زیر بدست می‌آید:

$$E_{\text{out}}(g^{(\mathcal{D})}) = \mathbb{E}_x\left[(f(x) - g^{(\mathcal{D})}(x))^۲\right]$$

در واقع در این رابطه، مقدار میانگین خطای داده‌های خارج را با استفاده از اختلاف $g(\mathcal{D})$ با تابع مولد داده‌ها محاسبه می‌کنیم. این مقدار به ما در بررسی عملکرد مدل کمک می‌کند و کم بودن آن بیانگر قابلیت مدل در تعمیم بر روی داده‌های دیده نشده است، و در صورتی که مقدار آن زیاد باشد نتیجه می‌گیریم که مدل عملکرد مناسبی بر روی داده‌های دیده نشده ندارد. اگر میانگین این عبارت را نسبت به همه‌ی $\mathcal{D}$های ممکن محاسبه کنیم، خواهیم داشت:

$$\mathbb{E}_{\mathcal{D}} E_{\text{out}}(g^{(\mathcal{D})}) = \text{bias}^۲ + \text{var}$$

در این صورت بدون در نظر گرفتن هیچ $\mathcal{D}$ خاص می‌توانیم عملکرد مدل را مورد بررسی قرار دهیم. حال اگر در داده‌هایمان نویز داشته باشیم، یعنی بجای دسترسی به داده‌هایی که از تابع $f(x)$ آمده‌اند، به داده‌هایی دسترسی داشته باشیم که از تابع $y(x) = f(x) + \varepsilon$ آمده باشند، می‌توانیم بگوییم:

$$E_{\text{out}}(g^{(\mathcal{D})}) = \mathbb{E}_{x,y}\left[(y(x) - g^{(\mathcal{D})}(x))^۲\right]$$

در واقع در این حالت فرض می‌کنیم که مانند قبل تابع $f(x)$ای وجود دارد که به دنبال یافتن آن هستیم، اما هنگام تولید داده‌ها، مقداری نویز وجود داشته که مقادیر خروج را از مقادیر واقعی آنها متفاوت کرده است و در نتیجه داده‌های خروج به شکل $y(x)$ هستند و نه $f(x)$. در چنین شرایطی، به دنبال یافتن تابعی هستیم که همچنان خود $f(x)$ (و نه $y(x)$) را به میزان خوبی تخمین بزند. متغیر تصادفی ε، همان نویز بوده و میانگین و واریانس آن را در این مثال به ترتیب صفر و $\sigma^۲$ در نظر می‌گیریم. با توجه به توضیحات و فرض داده شده، نشان دهید:

$$\mathbb{E}_{\mathcal{D}} E_{\text{out}}(g^{(\mathcal{D})}) = \text{bias}^۲ + \text{var} + \sigma^۲$$

۲. (نمره) فرض کنید در مسئله‌ی رگرسیون مقداری نویز به صورت زیر داشته باشیم:

$$X \in \mathbb{R}^{n\times d}, \quad w \in \mathbb{R}^d, \quad y \in \mathbb{R}^n, \quad \varepsilon \sim \mathcal{N}(۰, I_n)$$

$$y + \varepsilon = Xw$$

برای سادگی مقدار $\tilde{y} \in \mathbb{R}^n$ را به صورت $\tilde{y} = y + \varepsilon$ تعریف می‌کنیم. داده‌های ما به گونه‌ای باشند که برای مقادیر بدون نویز y رابطه‌ی $y = Xw$ برقرار است. اکنون با فرض اینکه تنها مقادیر X و $\tilde{y}$ برای آموزش مدل در دسترس باشند، به پرسش‌های زیر پاسخ دهید.

آ. پاسخ مدل را برای نقطه‌ی فرض $z \in \mathbb{R}^d$ با $\tilde{f}(z)$ نمایش می‌دهیم. عبارت $\tilde{f}(z)$ بر حسب z چیست؟

ب. می‌خواهیم امید ریاضی خطای مدل یعنی $\mathbb{E}[(\tilde{f}(z) - \tilde{y})^۲]$ محاسبه کنیم. تجزیه‌ی بایاس–واریانس این عبارت را انجام دهید و مشخص کنید هر کدام از بخش‌های بدست آمده نشان‌دهنده‌ی چه چیزی هستند.

پ. بایاس مدل را محاسبه کنید.

ت. نشان دهید که واریانس مدل برابر با $z^\top(X^\top X)^{-۱}z$ می‌باشد.

ث. چرا در اینجا برای هر نقطه‌ی تست دلخواه مثل z واریانس حداقل برابر با بایاس می‌باشد؟

۳. (نمره) دو عضو $a, b \in \mathbb{R}^۲$ را درنظر بگیرید.

۱. نشان دهید فضای ویژگی‌ای که کرنل $k_۱(a,b) = (a^\top b)^۲$ را تعریف می‌کند به شکل زیر است:

$$\phi([x_۱, x_۲]^\top) = \begin{pmatrix} x_۱^۲ \\ \sqrt{۲}x_۱x_۲ \\ x_۲^۲ \end{pmatrix}$$

۲. نشان دهید فضای ویژگی‌ای که کرنل $k_۲(a,b) = (۱ + a^\top b)^۲$ را تعریف می‌کند به شکل زیر است:

$$\phi([x_۱, x_۲]^\top) = \begin{pmatrix} x_۱^۲ \\ \sqrt{۲}x_۱x_۲ \\ x_۲^۲ \\ \sqrt{۲}x_۱ \\ \sqrt{۲}x_۲ \\ ۱ \end{pmatrix}$$

۴. (نمره) فرض کنید مجموعه‌داده‌ای به شکل

$$\{(۱, ۰), (\theta, ۱), (-۱, ۰)\}$$

برای مسئله‌ی رگرسیون در اختیار داریم که

$$۰ \leq \theta.$$

دو مدل

$$h_۰(x) = b$$

و

$$h_۱(x) = ax + b$$

را در نظر بگیرید. میزان خطای LOOCV را با استفاده از معیار خطای میانگین مربعات خطا برای هر کدام از این دو مدل محاسبه نمایید و مشخص کنید که به ازای چه مقداری از θ میزان خطای دو مدل ارائه‌شده برابر می‌شود.

۵. (نمره) قانون به‌روزرسانی بردار وزن در پرسپترون را در نظر بگیرید:

if $x^{(t)}$ is misclassified then:
$\boldsymbol{\omega}^{(t+1)} = \boldsymbol{\omega}^{(t)} + \eta\, y^{(t)}\boldsymbol{x}^{(t)}$

نشان دهید که در دسته‌بند پرسپترون، بردار وزن $\boldsymbol{\omega}$ را می‌توان به صورت ترکیب خطی داده‌ها نوشت:

$$\boldsymbol{\omega} = \sum_{i=1}^{N} \alpha_i \boldsymbol{x}_i$$

ضرایب α_i را در این ترکیب خطی مشخص کنید.