دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

مدرس: دکتر سید ابوالفضل مطهری

تمرین اول

۱. برای هر یک از موارد زیر، درستی یا نادرستی گزاره را با ذکر یک دلیل کوتاه مشخص کنید.

(آ) اگر مدل رگرسیون خطی با استفاده از ۸۰ درصد داده‌ها آموزش داده شود، نسبت به حالتی که از کل داده‌ها برای آموزش آن استفاده شود، بایاس کمتری خواهد داشت.

(ب) شرایطی وجود دارد که تحت آن، هم بایاس و هم واریانس متناظر یک مدل، بالا باشد.

(ج) اگر دقت یک مدل روی داده‌های آموزشی خوب و روی داده‌های آزمون (تست) کم باشد، استفاده از مدلی پیچیده‌تر می‌تواند به رفع مشکل کمک کند.

۲. در مسئله رگرسیون، فرض کنید مقادیر y با استفاده از یک تابع چندجمله‌ای درجه ۵ از مقادیر x متناظرشان به دست آمده‌اند. با در نظر گرفتن این امر، سه نوع مدل را در نظر گرفته‌ایم تا رابطه میان x و y را به صورت تخمینی مشخص کنیم. برای هر یک از مدل‌های زیر، با ذکر دلیل مختصر مشخص کنید بایاس و واریانس تخمین‌های متناظرشان روی داده‌ها (با توجه به مدل واقعی) را با عبارت‌های «کم» یا «زیاد» مشخص کنید.

(آ) رگرسیون خطی

(ب) چندجمله‌ای درجه ۵

(ج) چندجمله‌ای درجه ۱۲

۳. در مسئله رگرسیون خطی با روش کمترین مربعات، می‌دانیم که می‌توان ضرایب مدل را از طریق تخمین‌گر $\hat{\theta} = (X^TX)^{-1}X^TY$ به دست آورد. تحت چه شرایطی، $\hat{\theta}$ تخمین‌گر بیشینه درست‌نمایی نیز خواهد بود؟

۴. رگرسیون خطی را در حالت ساده در نظر بگیرید که در آن قصد پیش‌بینی Y بر حسب X را داریم. ثابت کنید آماره‌ی R^2 برابر است با مجذور همبستگی X و Y. برای سادگی می‌توانید فرض کنید $\bar{x} = \bar{y} = 0$.

۵. یک مسئله رگرسیون را در نظر بگیرید که در آن می‌خواهیم مقدار y را از روی یک مقدار حقیقی مانند x پیش‌بینی کنیم. تعدادی نمونه مانند $(x_i, y_i)_{i=1}^{n}$ داریم ($n \geqslant 3$). دو مدل زیر را برای مسئله رگرسیون خطی معمولی در نظر گرفته‌ایم.

(آ) $y_i = w_0 + w_1x_i + \epsilon_i$

(ب) $y_i = w_0 + w_1x_i + w_2x_i^2 + \epsilon_i$

توجه کنید که مقادیر ϵ_i مستقل و با توزیع یکسان (i.i.d) هستند و از توزیع گوسی با میانگین صفر می‌آیند. به سوالات زیر پاسخ دهید.

(آ) در مدل (ب)، رابطه‌ای برای تخمین w_2 به دست آورید. فرض کنید مقادیر w_0 و w_1 را می‌دانیم.

(ب) کدام یک از مدل‌ها دقت بیشتری روی داده‌های آموزشی خواهد داشت؟ (با ذکر دلیل)

(ج) کدام یک از مدل‌ها دقت بیشتری روی داده‌های آزمون (تست) خواهد داشت؟ (با ذکر دلیل)

۶. **(رگرسیون خطی وزن‌دار)**: در مسئله رگرسیون خطی، قصد داریم به نمونه‌های آموزشی، وزن‌های متفاوتی نسبت دهیم. به بیان دقیق‌تر، می‌خواهیم مقدار $J(\theta)$ را کمینه کنیم که به صورت زیر تعریف می‌گردد:

$$J(\theta) = \frac{۱}{۲} \sum_{i=۱}^{m} w^{(i)} (\theta^T x^{(i)} - y^{(i)})^{۲}.$$

(آ) نشان دهید ماتریس W موجود است؛ به طوری که داریم:

$$J(\theta) = (X\theta - y)^T W (X\theta - y)$$

(ب) با محاسبه $\nabla_\theta J(\theta)$ و برابر قرار دادن آن با صفر، مقدار θـ ای را که $J(\theta)$ را کمینه می‌کند، بیابید. (توجه: در حالتی که همه وزن‌ها یکسان باشند، می‌دانیم $\theta^* = (X^T X)^{-۱} X^T y$. جواب‌تان برای این قسمت باید یک فرم بسته باشد که تابعی از X، W و y است).

(ج) فرض کنید مجموعه داده $\{(x^{(i)}, y^{(i)}) : i = ۱, ۲, \ldots, m\}$ شامل m نمونه مستقل داده شده است. قصد داریم $y^{(i)}$ـ ها را به گونه‌ای مدل کنیم که گویی از توزیع‌های شرطی با سطوح مختلفی از واریانس گرفته شده‌اند. به طور مشخص، فرض کنید داریم:

$$p(y^{(i)} | x^{(i)}; \theta) = \frac{۱}{\sqrt{۲\pi}\sigma^{(i)}} \exp(-\frac{(y^{(i)} - \theta^T x^{(i)})^{۲}}{۲(\sigma^{(i)})^{۲}})$$

به بیان دیگر، $y^{(i)}$ از یک توزیع گوسی با میانگین $\theta^T x^{(i)}$ و واریانس $(\sigma^{(i)})^{۲}$ می‌آید؛ $\sigma^{(i)}$ـ ها ثابت هستند و مقدارشان مشخص است. نشان دهید که یافتن تخمین بیشینه درست‌نمایی برای θ، معادل است با حل یک مسئله رگرسیون خطی وزن‌دار. به طور مشخص مقادیر $w^{(i)}$ـ ها را بر حسب $\sigma^{(i)}$ـ ها به دست آورید.