به نام خدا

پاسخ بخش تئوری تمرین اول درس یادگیری ماشین

استاد درس: دکتر یاسمن برشبان

علی افشار دگرسی ۱۴۰۴۱۲۲۶۱۵۱۰۱

سوال اول: : بایاس -واریانس

الف) با افزایش تعداد داده های آموزشی، واریانس و بایاس مدل یاد گرفته شده چگونه تغییر میکند ؟

پاسخ: با افزایش داده‌های آموزشی، واریانس و بایاس مدل ما کاهش پیدا می‌کند. علت این اتفاق نیز آن است که مدل ما با دیدن نمونه‌های بیشتر، generalization یا تعمیم پذیری بیشتری خواهد داشت و در روی داده‌های validation خطای کمتری را تجربه خواهد کرد. به عبارتی دیگر، با افزایش داده‌های آموزشی، پیچیدگی مدل قابل درک تر شده و بایاس کاهش میابد، همچنین افزایش داده باعث حساسیت کمتر مدل به داده شده و واریانس کاهش میابد.

ب) در مساله رگرسیون، از تابع درجه 5 برای تخمین استفاده شده است، اما خطای آموزش قابل توجهی داریم. درجه تخمین را باید چگونه تغییر دهیم؟

پاسخ: با توجه به اینکه در بخش آموزش خطای قابل توجهی داریم، پس دچار under-fitting شده ایم. به این علت نیاز هست که درجه تخمین خود را افزایش دهیم تا به کمک آن روی داده‌های آموزشی خود خطای کمتری را تجربه کنیم.

سوال دوم: دسته بندی خطی

الف) مشکل استفاده از تابع هزینه SSE در مسأله دسته‌بندی چیست؟

پاسخ: معادله تابع هزینه SSE به صورت زیر است:

$$J(w) = \sum_{i=1}^{N} (w^T x^{(i)} - y^{(i)})^2$$

با توجه به تابع هزینه، در مساله دسته‌بندی دو کلاسه، y های واقعی ما معمولا بین دو عدد {۱+،۱-} تعریف می‌شوند، با توجه به این تعریف اگر فاصله یک نمونه از کلاس ۱+ از خط پیشبینی شده ما بیشتر از **صفر** باشد ولی برابر **یک** نباشد(بیشتر یا کمتر از یک باشد) در این صورت برخلاف اینکه این نمونه بالای خط است(فاصله از خط مثبت است) این تابع برای این نمونه خطایی بیشتر از **صفر** در نظر میگیرد که اشتباه است.

ب)برتری تابع هزینه روش دسته بندی پرسپترون نسبت به تابع هزینه ای که تنها تعداد نمونه های غلط دسته بندی شده را در نظر میگیرد چیست؟

پاسخ:

- تابع هزینه در روش دسته‌بندی پرسپترون در تمام نقاط خود **مشتق پذیر** است ولی تعداد نمونه های غلط دسته‌بندی شده در بعضی نقاط(در هنگام تغییر مقدار به صورت پله‌ای تغییر می کند) مشتق پذیر نیست. مشتق پذیر بودن یا نبودن تابع هزینه در تمام نقاط، برای حل مسائل بهینه سازی در روش Gradient Descent که در هر مرحله ضریبی از گرادیان(همان مشتق) را از پارامترهای مدل کم می کند، لازم می‌باشد.

- تابع هزینه در روش دسته‌بندی پرسپترون، مقدار خطا، که برابر مقدار فاصله نمونه دچار خطا، از خط پیش‌بینی کننده ما است را برای هر نمونه محاسبه میکند، که داده بهتری برای اندازه گیری **شدت خطا**(هرچه نمونه از خط دورتر باشد خطای آن بیشتر است) به نسبت فقط تعداد نمونه‌های اشتباه دسته‌بندی شده می‌باشد. به بیان دیگر میتوان گفت که تابع تعداد نمونه های غلط دسته‌بندی شده زیر مجموعه تابع هزینه پرسپترون است.

سوال سوم: LDA

فرض کنید داده های مساله، نقاط زیر در فضای دو بعدی باشند. راستای بهینه را با استفاده از الگوریتم LDA محاسبه کنید.

$$C_1 = \{(1,1), (1,2), (2,1), (2,4), (3,1), (3,3)\}$$

$$C_2 = \{(2,2), (3,4), (4,2), (5,1), (5,4), (5,5)\}$$

پاسخ:

$$\mu_1 = \frac{1}{6}\begin{bmatrix} \overbrace{1+1+2+2+3+3}^{12} \\ \underbrace{1+2+1+4+1+3}_{12} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\mu_2 = \frac{1}{6}\begin{bmatrix} \overbrace{2+3+4+5+5+5}^{24} \\ \underbrace{2+4+2+1+4+5}_{18} \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} \overbrace{1-2}^{-1} \\ \underbrace{1-2}_{-1} \end{bmatrix} \begin{bmatrix} \overbrace{1-2}^{-1} & \overbrace{1-2}^{-1} \end{bmatrix} + \begin{bmatrix} \overbrace{1-2}^{-1} \\ \underbrace{2-2}_{0} \end{bmatrix} \begin{bmatrix} \overbrace{1-2}^{-1} & \overbrace{2-2}^{0} \end{bmatrix} + \begin{bmatrix} \overbrace{2-2}^{0} \\ \underbrace{1-2}_{-1} \end{bmatrix} \begin{bmatrix} \overbrace{2-2}^{0} & \overbrace{1-2}^{-1} \end{bmatrix} +$$

$$\begin{bmatrix}\overbrace{2-2}^{0}\\ \underbrace{4-2}_{2}\end{bmatrix}\begin{bmatrix}\overbrace{2-2}^{0} & \overbrace{4-2}^{2}\end{bmatrix}+\begin{bmatrix}\overbrace{3-2}^{1}\\ \underbrace{1-2}_{-1}\end{bmatrix}\begin{bmatrix}\overbrace{3-2}^{1} & \overbrace{1-2}^{-1}\end{bmatrix}+\begin{bmatrix}\overbrace{3-2}^{1}\\ \underbrace{3-2}_{1}\end{bmatrix}\begin{bmatrix}\overbrace{3-2}^{1} & \overbrace{3-2}^{1}\end{bmatrix}$$

$$S_1=\begin{bmatrix}\overbrace{1+1+0+0+1+1}^{4} & \overbrace{1+0+0+0+(-1)+1}^{1}\\ \underbrace{1+0+0+0+(-1)+1}_{1} & \underbrace{1+0+1+4+1+1}_{8}\end{bmatrix}$$

$$S_2=\begin{bmatrix}\overbrace{2-4}^{-2}\\ \underbrace{2-3}_{-1}\end{bmatrix}\begin{bmatrix}\overbrace{2-4}^{-2} & \overbrace{2-3}^{-1}\end{bmatrix}+\begin{bmatrix}\overbrace{3-4}^{-1}\\ \underbrace{4-3}_{1}\end{bmatrix}\begin{bmatrix}\overbrace{3-4}^{-1} & \overbrace{4-3}^{1}\end{bmatrix}+\begin{bmatrix}\overbrace{4-4}^{0}\\ \underbrace{2-3}_{-1}\end{bmatrix}\begin{bmatrix}\overbrace{4-4}^{0} & \overbrace{2-3}^{-1}\end{bmatrix}+$$

$$\begin{bmatrix}\overbrace{5-4}^{1}\\ \underbrace{1-3}_{-2}\end{bmatrix}\begin{bmatrix}\overbrace{5-4}^{1} & \overbrace{1-3}^{-2}\end{bmatrix}+\begin{bmatrix}\overbrace{5-4}^{1}\\ \underbrace{4-3}_{1}\end{bmatrix}\begin{bmatrix}\overbrace{5-4}^{1} & \overbrace{4-3}^{1}\end{bmatrix}+\begin{bmatrix}\overbrace{5-4}^{1}\\ \underbrace{5-3}_{2}\end{bmatrix}\begin{bmatrix}\overbrace{5-4}^{1} & \overbrace{5-3}^{2}\end{bmatrix}$$

$$S_2=\begin{bmatrix}\overbrace{4+1+0+1+1+1}^{8} & \overbrace{2+(-1)+0+(-2)+1+2}^{2}\\ \underbrace{2+(-1)+0+(-2)+1+2}_{2} & \underbrace{1+1+1+4+1+4}_{12}\end{bmatrix}$$

$$S_W=S_1+S_2=\begin{bmatrix}12 & 3\\ 3 & 20\end{bmatrix}\qquad \det(S_W)=(12\times 20)-(3\times 3)=231$$

$$S_W^{-1}=\frac{1}{\underbrace{\det(S_W)}_{231}}\begin{bmatrix}20 & -3\\ -3 & 12\end{bmatrix}$$

$$\mathbf{w}=S_W^{-1}(\mu_1-\mu_2)=\frac{1}{231}\begin{bmatrix}20 & -3\\ -3 & 12\end{bmatrix}\begin{bmatrix}\overbrace{2-4}^{-2}\\ \underbrace{2-3}_{-1}\end{bmatrix}=\frac{1}{231}\begin{bmatrix}\overbrace{20\times(-2)}^{-40}+\overbrace{(-3)\times(-1)}^{3}\\ \underbrace{(-3)\times(-2)}_{6}+\underbrace{12\times(-1)}_{-12}\end{bmatrix}$$

$$\boxed{\mathbf{w} = \frac{1}{231}\begin{bmatrix}-37\\-6\end{bmatrix}}$$