رضا قربانی ۴۰۳۲۰۹۵۹۵ در تمرین ۱ ML دوم

سوال 1)

الف) چرا از تابع softmax اغلب برای مسائل دسته‌بندی استفاده می‌شود؟

تابع softmax اغلب در مسائل دسته‌بندی چندکلاسه استفاده می‌شود، زیرا خروجی آن احتمال نسبی هر کلاس را نشان می‌دهد. این تابع اطمینان حاصل می‌کند که مجموع احتمال‌ها برابر 1 باشد، بنابراین به مدل کمک می‌کند که به صورت بهینه بین چندین کلاس تصمیم‌گیری کند. هرچه مقدار پیش‌بینی برای یک کلاس بالاتر باشد، احتمال تخصیص آن کلاس بیشتر است.

ب) بالا بودن واریانس در مدل چه معنایی دارد؟ یک روش ممکن برای کاهش واریانس در مدل خود بیان کنید.

بالا بودن واریانس در یک مدل به معنای آن است که مدل بر داده‌های آموزشی بیش‌ازحد پیچیده‌شده و توانایی تعمیم به داده‌های جدید را ندارد (پدیده overfitting) برای کاهش واریانس، می‌توان از روش‌های کاهش پیچیدگی مانند کاهش تعداد ویژگی‌ها، استفاده از تنظیم‌کننده‌ها مثل L2 یا Ridge regularization، یا استفاده از داده‌های بیشتر برای آموزش مدل بهره برد.

پ) چرا در حالتی که تمام ویژگی‌ها تا حد خوبی با خروجی مرتبط هستند، رگرسیون Ridge به رگرسیون Lasso ترجیح داده می‌شود؟

در چنین حالتی که تمام ویژگی‌ها دارای اهمیت و همبستگی با خروجی هستند، رگرسیون Ridge مناسب‌تر است زیرا این روش به جای حذف کامل ویژگی‌ها (که Lasso انجام می‌دهد)، ضرایب ویژگی‌ها را کاهش می‌دهد. در نتیجه، اطلاعات هیچ ویژگی حذف نمی‌شود و تمام آن‌ها به مدل کمک می‌کنند. اما Lassoویژگی‌های کم‌اهمیت را به طور کامل حذف می‌کند که ممکن است منجر به از دست رفتن اطلاعات مفید شود.

ت) چگونه رگرولاریزیشن L2 در classifier های خطی بر روی تعادل بایاس-واریانس تأثیر می‌گذارد؟

رگرولاریزیشن(L2 که همان Ridge است) در دسته‌بندهای خطی باعث کاهش واریانس می‌شود، زیرا ضرایب ویژگی‌ها را کوچک می‌کند و مدل را ساده‌تر می‌سازد. این به کاهش پدیده overfittingکمک می‌کند. با این حال، ممکن است کمی بایاس را افزایش دهد زیرا مدل به سمت حالت کم‌پیچیدگی حرکت می‌کند. در کل، L2 به تعادل بهتری بین بایاس و واریانس کمک می‌کند، با کاهش واریانس به قیمت کمی افزایش بایاس.

سوال2)

الف) برای محاسبه بهترین w_j باید کمینه سازی MSE را انجام بدهیم

$$\min_{w} \| y - w_j x_j \|^2$$

حال مشتق می گیریم:

$$\frac{\partial}{\partial w_j} \| y - w_j x_j \|^2 = -2 x_j^T (y - w_j x_j)$$

مشتق را برابر صفر قرار می دهیم

$$x_j^T y = w_j x_j^T x_j$$

$$w_j = \frac{x_j^T x_j}{x_j^T y}$$

ب) در این قسمت فرض می کنیم که ویژگی های مختلف مستقل از یکدیگرند

از آنجایی که ویژگی ها مستقل هستند این مسئله به n مسئله تقسیم می شود

که هر کدام به شکل زیر هستند

$$w_j = \frac{x_j^T y}{x_j^T x_j}$$

ب) در اینجا ما یه ترم bias هم به معادله‌مون اضافه میشه

و ما باید MSE رو به شکل زیر minimize کنیم

$$\min \| y - (w_j x_j + w_0) \|^2$$

نسبت به w_0 مشتق می‌گیریم

$$\frac{\partial}{\partial w_0} \| y - w_0 - w_j x_j \|^2 = -2 \sum (y - w_0 - w_j x_j)$$

مشتق رو برابر صفر قرار میدیم

$$w_0 = \frac{1}{N} \sum y - w_j \frac{1}{N} \sum x_j$$

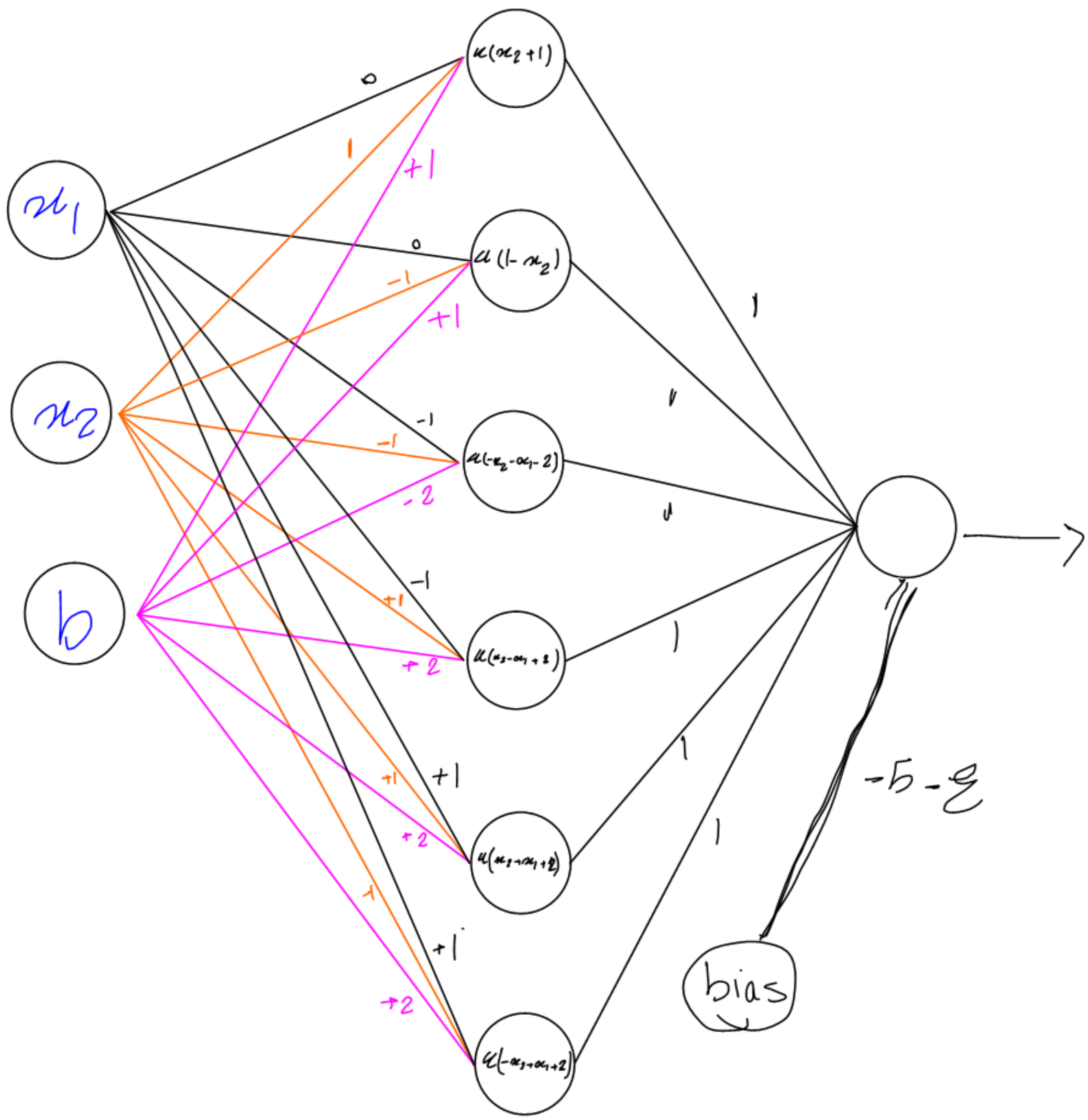
همین کار را به صورت مشابه برای x_j انجام میدیم و خواهیم داشت

$$w_j = \frac{x_j^T (y - w_0)}{x_j^T x_j}$$

سوال 3)

معادله خط های ۶ ضلعی به ترتیب زیر است .

$$
\begin{cases}
x_2+1=0 \longrightarrow x_2+1>0 \\
x_2-1=0 \longrightarrow 1-x_2>0 \\
x_2+x_1-2=0 \longrightarrow -x_1-x_2+2>0 \\
x_2-x_1+2=0 \longrightarrow x_2-x_1+2>0 \\
x_2+x_1+2=0 \longrightarrow x_2+x_1+2>0 \\
x_2-x_1-2=0 \longrightarrow -x_2+x_1+2>
\end{cases}
$$



سوال 4)

الف) مشتق softmax را بر حسب z_i $(i=k)$ حساب می‌کنیم:

$$\frac{\partial \hat{y}_i}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}\right)$$

$$= \frac{e^{z_i} \cdot \left(\sum e^{z_j}\right) - e^{z_i} e^{z_i}}{\left(\sum e^{z_i}\right)^2}$$

$$= \frac{e^{z_i}}{\sum e^{z_j}}\left(1 - \frac{e^{z_i}}{\sum e^{z_j}}\right)$$

$$\boxed{\frac{\partial \hat{y}_i}{\partial z_i} = \hat{y}_i (1 - \hat{y}_i)}$$

حالت $(i \neq k)$

$$\frac{\partial \hat{y}_i}{\partial z_k} = -\frac{e^{z_i} \cdot e^{z_k}}{\left(\sum e^{z_j}\right)^2}$$

$$= -\hat{y}_i \hat{y}_k$$

ج) از قاعده مشتق زنجیره ای برای محاسبه این مشتق استفاده می کنیم

$$\frac{\partial L}{\partial z_i} = \sum_{i=1}^{n} \frac{\partial L}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_k}$$

$$\frac{\partial L}{\partial z_i} = -\sum y_i \cdot \frac{1}{\hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_i}$$

با توجه به مشتق های تابع softmax درحالت ها z_i $(i=k)$ و $(i \neq k)$ داریم :

$$\frac{\partial L}{\partial z_i} = \hat{y}_i - y_i$$