

به نام خدا

سینا کریمی

9931050

تمرین دوم

سوال 1)

الف)

ابعاد ماتریس حاصل برابر با مقدار زیر خواهد بود:

$$\left(\frac{N-F}{S}\right) + 1 = \frac{5-3}{2} + 1 = 2$$

حال ضرب ماتریس ها و مقادیر آنها:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = 5$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = 3$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = 4$$

$$\begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = 5$$

$$\begin{bmatrix} 5 & 3 \\ 4 & 5 \end{bmatrix}$$

(ب)

$$0.5 \left( \begin{bmatrix} 5 & 3 \\ 4 & 5 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \right) = \begin{bmatrix} 1 & 1.5 \\ 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 0 & 0 \end{bmatrix} * 1 + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} * 1.5 + \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} * 2 + \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \end{bmatrix} * 1$$

$$= \begin{bmatrix} 4.5 & 4 & 2 \\ 6.5 & 5.5 & 3 \\ 3 & 5 & 5.5 \end{bmatrix}$$

$$W = W - a * GradW$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} - 0.5 * \begin{bmatrix} 4.5 & 4 & 2 \\ 6.5 & 5.5 & 3 \\ 3 & 5 & 5.5 \end{bmatrix} = \begin{bmatrix} -1.25 & -2 & -0.5 \\ -3.75 & -1.75 & -1.5 \\ -0.5 & -2.5 & -1.75 \end{bmatrix}$$

سوال 2)

(الف)

سناریو اول نشان دهنده overfitting است به دلیل تفاوت زیاد در درصد داده های آموزشی و تستی نشان میدهد مدل به خوبی با داده های آموزشی وفق دارد ولی در تست اینگونه نیست

سناریو دوم نشاندهنده Underfitting است چرا که هردو داده آموزشی و تست دارای دقت پایینی هستند

سناریو سوم نشان دهنده overfitting است به دلیل تفاوت زیاد در درصد داده های آموزشی و تستی نشان میدهد مدل به خوبی با داده های آموزشی وفق دارد ولی در تست اینگونه نیست و توانایی عمومی سازی ندارد

(ب)

سناریوی 1 (بیش برازش)

استفاده از تکنیک های regularization مانند L1 و L2

کاهش تعداد نوروها یا کاهش پیچیدگی مدل

استفاده از dropout برای کاهش وابستگی نوروها به یکدیگر

سناریوی 2 (پیش برازش)

افزایش تعداد نوروها در لایه پنهان برای افزایش ظرفیت مدل

استفاده از مدل های پیچیده تر

افزایش تعداد دوره های آموزشی (epochs) تا مدل به خوبی آموزش داده شود

سناریوی 3 (بیش برازش)

استفاده از تکنیک های regularization مانند L1 و L2

استفاده از dropout برای کاهش وابستگی نوروها به یکدیگر

کاهش تعداد نوروها در لایه پنهان یا استفاده از معماری های ساده تر

سوال 3)

الف و ب)

Dropout: روش برای جلوگیری از overfitting می باشد. وقتی یک لایه fully connected تعداد زیادی نورو دارد، شرایطی میتواند پیش بیاید که در آن تعدادی از نورونی ها ویژگی های یکسان

یا شبیه به هم را از داده ها استخراج میکنند که میتواند منجر به overfitting بشود. با استفاده از dropout درصدی از نورون ها به صورت تصادفی خاموش میشوند. این روش فقط در لایه های پنهان استفاده میشود

Regularization: یک تکنیک برای رفع overfitting است که در آن یک penalty به loss function اضافه میشود و مدل را از اهمیت دادن به بعضی از ویژگی ها و ضرایب برحضر میدارد (ج) تفاوت L1 و L2 در این است که L2 از مربع وزن ها استفاده میکند.

زمانی از L1 استفاده میکنیم که در آموزش مدل خود داریم feature selection انجام میدهیم و تنها میخواهیم از بخشی از feature های داده ها که مهم هستند استفاده کنیم

زمانی از L2 استفاده میکنیم که میخواهیم ضرایب را کاهش دهیم ولی همچنان به تمام feature ها نیاز داریم و نمیخواهیم آنها را حذف کنیم و feature ها به مربوط هستند (multicollinearity)

سوال 4)

الف) SOM ها یا self organizing map ها نوعی شبکه عصبی هستند که در unsupervised learning و data visualization استفاده میشوند و هدف اصلی آنها کاهش ابعاد داده ها در عین نگهداشتن رابطه میان آنها است (در clustering استفاده میشود). شامل یک لایه ورودی و یک لایه خروجی است

ب) میتوان وزن ها را صورت تصادفی مقداردهی کرد که ساده ترین راهکار است ولی منجر به همگرایی آهسته تر میشود. میتوان به صورت خطی مقداردهی کرد که شروع ساختار یافته تری خواهیم داشت و به همگرایی سریع تری منجر میشود. اهمیت وزن دهی:

1- در سرعت همگرایی تاثیر دارد

2- در پایداری نگاشت ها تاثیر دارد و میتواند از دوری از کردن از مینیمم محلی کمک کند

3- کیفیت خوشه بندی را تحت تاثیر قرار میدهد به طوری که به خوشه بندی بهتر و نمایش بهتری از داده ها منجر میشود

ج) اگر تعداد گره ها کم باشد نمیتواند به درستی روابط داده ها را بیابد و منجر به underfitting میشود. در مقابل اگر تعداد گره ها زیاد باشد جزئیات بیشتری را بدست میآورد ولی میتواند منجر به overfitting میشود. در کنار این موارد میتواند زمان train کردن تاثیر بگذارد که هرچه تعداد گره ها بیشتر آپدیت های بیشتری باید رخ بدهد و حافظه بیشتری استفاده میکند

سوال 5)

(الف)

1. ساخت مراکز به صورت تصادفی
2. تعیین فاصله هر داده با مراکز
3. انتساب یک داده به نزدیک ترین مرکز
4. تعیین مرکز جدید با توجه به داده اضافه شده
5. تکرار مراحل 3 و 4

(ب)

از نوع unsupervised هستند از آنجایی که داده ها label ندارند

(ج)

پیش برآزش (underfitting)

(د)

میان خوشه ای (intra cluster) را کم و درون خوشه ای (inter cluster) را زیاد میکنیم. هرچه میان خوشه ای کمتر، شباهت و ربط داده های درون یک خوشه بیشتر و هرچه درون خوشه ای بیشتر، overlap خوشه ها کمتر و متمایز تر هستند

سوال 6)

شبکه عصبی convolution یک کلاس از شبکه های عصبی است که برای پردازش موثر داده های grid شکل مانند عکس ها بکار میرود. این شبکه دارای چندین لایه است که به ترتیب بر روی پردازش داده کار میکنند:

- 1- لایه convolution که با استفاده از filter ها ویژگی های ورودی را استخراج میکند
- 2- لایه pooling که ورودی را downscale میکند تا پیچیدگی های محاسباتی را کاهش دهد
- 3- لایه fully connected که وظیفه پیشبینی را برعهده دارند و خروجی را تولید میکنند

(ب)

کاربردهای آن بیشتر در image classification و image recognition است و مزیت های آن عبارتند از:

- 1- نیاز به supervision ندارند
- 2- استخراج خودکار featureها
- 3- دقت بالا و کاهش محاسبات
- 4- قابلیت هندل کردن dataset های بزرگ

(ج)

از لایه pooling برای downscale کردن خروجی لایه convolution و در نتیجه کاهش حجم محاسبات استفاده میشود و به دو صورت max pooling و average pooling صورت میگیرد

(د)

از فیلترها برای استخراج ویژگی ها از dataset استفاده میشود و به صورت که فیلترها بر روی داده ورودی به اصطلاح لغزنده (slide) میشوند و حاصل ضرب بخشی از ورودی که فیلتر بر روی آن

است با فیلتر، خروجی لایه convolution را تشکیل میدهد و در ادامه با backprop وزن های  
موجود در فیلترها آپدیت میشوند.

سوال 7)

سوال ۷

اگر ضریب منظم سازی را برابر با  $\lambda$  در نظر بگیریم، با استفاده  
از نسب معکوس معادله زیر را خواهیم داشت:

$$W = (G^T G + \lambda I)^{-1} \cdot G^T Y$$

برای حالتی که یک خروجی داریم هدف ما به شرح زیر خواهد

بود:

$$\min \left\{ \frac{1}{2} [\|XW - Y\|^2 + \lambda \|W\|^2] \right\}$$

۱۴۰۱ / ۹ / ۱۰

$$\frac{1}{\gamma} [||xw - y||^2 + \lambda ||w||^2] = \frac{1}{\gamma} (xw - y)^T (xw - y) + \frac{1}{\gamma} w^T \lambda w$$

$$\Rightarrow \frac{1}{\gamma} (w^T x^T x w - w^T x^T y - y^T x w + y^T y) + \frac{1}{\gamma} \lambda w^T w$$

$$\Rightarrow \frac{1}{\gamma} (w^T x^T x w - 2y^T x w + y^T y) + \frac{1}{\gamma} \lambda w^T w$$

$$\text{مشتق نسبت به } w \Rightarrow w^T x^T x - y^T x + \lambda w = 0 \Rightarrow x^T x w - x^T y + \lambda w = 0$$

$$\Rightarrow (x^T x + \lambda I) w = x^T y$$

$$\Rightarrow w = (x^T x + \lambda I)^{-1} x^T y$$

شهادت آیت الله سید حسن مدرس (۱۳۱۶ هـ) و روز محسن

۱۴۰۱ / ۹ / ۱۱

$$w^{[1]} = (x^{[1]T} x + \lambda I)^{-1} x^{[1]T} y^{[1]}$$

$$w^{[2]} = (x^{[2]T} x + \lambda I)^{-1} x^{[2]T} y^{[2]}$$

$$\Rightarrow x_{m \times n}, w_{n \times 1}, y_{m \times 1}$$

شهادت میرزا کوچک خان جنگلی (۱۳۰۰ هـ)