



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

تمرین تئوری سری دوم

مبانی هوش محاسباتی

بهار 1404

استاد درس: دکتر سمانه حسینی

دستیار آموزشی: آرش آژند

موعد تحویل تکلیف: 3 اردیبهشت 1404

نکات تحویل تکلیف:

1. پاسخ های خود را حتما در سامانه یکتا آپلود کنید و از ارسال در بستر تلگرام خودداری کنید.
2. پاسخ ها ترجیحا تایپ شده و تمیز باشد که موجب کسر نمره به دلیل ناخوانایی نشود.
3. انجام تکلیف بصورت تک نفره است و در صورت مشاهده تقلب نمرات هم مبدأ و هم مقصد صفر میشود.
4. ساختار نامگذاری تکلیف باید بصورت روبرو باشد HWX_StudentID_LastName.pdf که X شماره تمرین، StudentID شماره دانشجویی و LastName نام خانوادگی میباشد.

آیدی تلگرام:

@arash_azhand

1- صحیح یا غلط بودن گزاره های زیر را مشخص کنید و برای جواب خود دلیل کامل بیاورید. (45 نمره)

- 1.1- افزایش حجم داده های آموزشی بطور کامل مشکل بیش برازش¹ را از بین می برد.
- 1.2- مدلی که دقت بالایی روی داده های آموزشی دارد اما دقت پایینی روی داده های اعتبارسنجی دارد، دچار بیش برازش شده است.
- 1.3- تنظیم L1 (Lasso) باعث صفر شدن برخی از وزن ها و کاهش ابعاد مدل می شود، در حالی که تنظیم L2 (Ridge) باعث کاهش مقدار وزن ها بدون صفر کردن آن ها می شود.
- 1.4- Inverted dropout هم در مرحله آموزش و هم در مرحله استنتاج استفاده می شود تا عملکرد مدل را بهبود دهد.
- 1.5- توقف زودهنگام² نوعی روش بهینه سازی ابر پارامترها محسوب می شود.
- 1.6- اگر اندازه مینی بچ را ۱ انتخاب کنیم، گرادیان نزولی با مینی بچ دقیقاً به گرادیان نزولی تصادفی³ تبدیل می شود.
- 1.7- استفاده از روش های نرمال سازی به فرار از بهینه های محلی در طول آموزش منجر میشود اما سرعت آموزش را کندتر میکنند.
- 1.8- برای تقسیم کردن مجموعه داده به زیر مجموعه های ارزیابی، آموزش و تست، نمونه برداری تصادفی از مجموعه داده به اندازه لازم برای هر زیر مجموعه همواره کفایت میکند.
- 1.9- در روش Exponentially Weighted Average، مدل، به تمام گرادیان های گذشته وزن مساوی میدهد.

2- به سوالات زیر پاسخ کامل دهید. (40 نمره)

- 2.1- چرا استفاده از تابع فعال ساز Tanh در لایه های میانی نسبت به Sigmoid ترجیح داده می شود؟

¹ Overfit

² Early Stopping

³ SGD

2.2- چرا توصیه می‌شود که برای تعیین مقدار نرخ آموزش در مدل‌های دارای Dropout، از نمودار هزینه بر حسب ایپاک استفاده نکنید؟

2.3- چگونه Exponentially Weighted Average می‌تواند به صاف کردن (Smooth کردن) یک سری زمانی پر از نویز کمک کند؟

2.4- چرا شبکه‌های عصبی عمیق نیاز به مقداردهی اولیه مناسب دارند؟ چه روشی برای این کار پیشنهاد می‌دهید؟

3- یک شبکه عصبی کم عمق fully connected با مشخصات زیر داده شده است:

- لایه ورودی: دو ویژگی X_1, X_2
- لایه پنهان: دو نورون با تابع فعال ساز sigmoid
- لایه خروجی: 1 نورون (دو کلاس) با فعال ساز sigmoid
- تابع هزینه: کراس انتروپی با رگیولاریزیشن L_2 ⁴
- از گرادیان کاهشی برای پیدا کردن وزن‌های بهینه استفاده میشود
- ضریب رگیولاریزیشن: $\lambda = 0.01$
- نرخ یادگیری: $\eta = 0.1$
- برای 2 ایپاک آموزش داده میشود

X1	X2	Class label (y)
0.5	1	0
1.5	2	1
2	1	0

جدول 1: سه نمونه ورودی در جدول داده شده است

پارامترهای اولیه شبکه:

$$b^{(1)} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix} \text{ و } w^{(1)} = \begin{bmatrix} 0.2 & -0.3 \\ 0.4 & 0.1 \end{bmatrix} \bullet$$

$$b^{(2)} = 0.05 \text{ و } w^{(2)} = \begin{bmatrix} 0.3 \\ -0.5 \end{bmatrix} \bullet$$

⁴ Ridge Regression

مدل را بصورت **vectorized** برای 2 ایپاک آموزش دهید و در هر ایپاک وزن های آپدیت شده را بدست آورید و حواستان به تمام نکات گفته شده برای شبکه باشد که موجب کسر نمره نشود.
(50 نمره)

4- فرض کنید میخواهید دمای روزانه یک شهر (بر حسب $^{\circ}\text{C}$) را با استفاده از EWA^5 با ضریب هموار ساز⁶ $\theta = 0.9$ محاسبه کنید. دمای ثبت شده در 5 روز گذشته به شرح زیر است:

- روز 1: 22°C
- روز 2: 24°C
- روز 3: 20°C
- روز 4: 19°C
- روز 5: 21°C

- مقدار EWA را برای هر روز حساب کنید، با فرض اینکه مقدار اولیه $V_0 = 20^{\circ}\text{C}$ باشد.
- انتخاب θ (مثلا مقایسه 0.5 و 0.9) چگونه حساسیت میانگین به داده های اخیر را تغییر میدهد؟ (40 نمره + 10 نمره امتیازی)

(امتیازی): برای مقایسه بین مقادیر مختلف θ ، کد آن را بنویسید و با پلات کردن روی نمودار آن را تحلیل کنید.

⁵ Exponentially Weighted Average

⁶ Decay Rate