



سلام بر تمام دانشجویان عزیز، چند نکته مهم:

1. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
2. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
3. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده از کنید. شکل ها، به طور واضح و در فرمت درست گزارش شود.
4. از بین سوالات **شبیه سازی** حتماً به **دو مورد** پاسخ داده شود. حداکثر تا نمره ۱۲۰ (۲۰ نمره امتیازی) لحاظ خواهد شد.
5. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله **تقلب** می باشد و کل نمره تمرین **صفر** می‌شود.
6. در صورت داشتن سوال، از طریق ایمیل salar.nouri@ut.ac.ir سوال خود را مطرح کنید.

1. سوالات زیر را بررسی کنید. (۲۰ نمره)

الف) در یک مسئله طبقه‌بندی، به دو صورت زیر عمل کرده‌ایم:

- 1) از ۴۰۰ داده برچسب‌گذاری شده موجود، ۲۰۰ تا برای آموزش و ۲۰۰ تا برای تست انتخاب کرده‌ایم.
 - 2) از ۴۰۰ داده برچسب‌گذاری شده موجود، ۳۸۰ نمونه برای آموزش و ۲۰ نمونه برای تست انتخاب کرده‌ایم.
- در حالت یک، به دقت ۸۰ درصد و در حالت دو به دقت ۹۰ درصد رسیده‌ایم. در مورد عملکرد هر دو و بهینه بودن آنها توضیح دهید.

ب) رابطه *inference* را با احتمال رخداد، با ذکر چند مثال بیان کنید.

2. در این سوال قصد داریم تا الگوریتم بهینه سازی گرادیان کاهشی را بررسی کنیم. (۲۰ نمره)

الف) ابتدا الگوریتم گرادیان کاهشی را توضیح دهید، سپس روابط آن را برای تابع هزینه زیر محاسبه نمایید.

$$J(\theta) = \frac{1}{2} \sum_{i=1}^q (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$h(x) = \tanh(w^T x + b)$$

ب) قانون به روزرسانی وزن‌ها و بایاس‌ها را به دست آورید. و اثر نرخ یادگیری را در قانون به روزرسانی وزن‌ها و بایاس توضیح دهید.

3. فرض کنید که (x_i, y_i) for $i = 1, \dots, k+1$ ، تعداد $k+1$ نقطه در صفحه xy می باشد، بدین صورت که هیچ کدام از نقاط، مقدار x یکسان ندارند. نشان دهید که یک چند جمله ای وجود دارد، که از این نقاط عبور می کند. (۲۰ نمره)

4. با در نظر گرفتن دو رابطه زیر برای مقادیر β_1, β_0 ، مقدار کواریانس مربوط به این دو پارامتر را حساب کنید. در چه حالتی این دو پارامتر مستقل می شوند. (۲۰ نمره)

$$\beta_1 = \frac{\sum_{i=1}^n (Y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \beta_0 = \bar{Y} - \beta_1 \bar{x}$$

5. با در نظر گرفتن جدول زیر، که مربوط به یک مسیله رگرسیون خطی ساده می باشد و در آن مقادیر Y چند نوع آلیاژ مختلف در دماهای X می باشند، سوالات مربوطه را بررسی کنید. (۲۰ نمره)

i	x_i	y_i
1	0.5	40
2	1.0	41
3	1.5	43
4	2.0	42
5	2.5	44
6	3.0	42
7	3.5	43
8	4.0	42

الف) مقدار پارامتر های $\beta_1, \beta_0, \sigma^2$ ، و همچنین واریانس مربوط به β_1, β_0 را محاسبه کنید.

ب) مقدار کوریلیشن مربوط به دو پارامتر β_1, β_0 را به دست آورید.

6. (شبه سازی) در این سوال، طبقه بندی طراحی کنید که بتوانیم، که ۲ کلاس متفاوت (دو تیم فوتبال منچستر یونایتد و چلسی) با استفاده از دیتاست داده شده، را تشخیص دهیم. جهت طبقه بندی، میتوانید میانگین رنگ در هر عکس را محاسبه نمایید، سپس بر اساس مقدار به دست آمده، با مقدار رنگ آبی و قرمز مقایسه نمایید. برای دیتاست داده شده، این طبقه بند را تست کنید و مقدار correct classification و ماتریس Confusion، Confidence را گزارش دهید. مقادیر accuracy, precision, recall را محاسبه کنید، و نتایج هر کدام را توضیح دهید. (۲۰ نمره)

7. (شبیه‌سازی) در این سوال می‌خواهیم، `overfitting` و `underfitting` را برای یک سری داده بررسی کنیم. ابتدا، به تعداد دلخواه نمونه کسینوسی تولید کرده، سپس این داده‌ها را با نویز گوسی میانگین صفر و واریانس دلخواه جمع کنید. حال سعی کنید که تابع درجه یک، چهار، هشت، دوازده، پانزده، بیست برای این داده‌ها برازش کنید. مقادیر `MSE` برای هر یک از موارد گزارش دهید. نمودار برازش شده بر این داده‌ها را رسم کنید. مشاهده خود را از نتایج به دست آمده شرح دهید. مقادیر بایاس و واریانس را ذکر کنید. (۲۰ نمره)

نکته: جهت تولید نمونه کسینوسی از $\cos(2 * \pi * x)$ و نویز از `random` در `numpy` استفاده کنید.

8. مفاهیم `accuracy`، `precision`، `recall` را با ذکر مثالی بیان کنید. همچنین در چه حالت‌های ممکن است یکی از این موارد جواب خوبی ندهد، که ملزم به استفاده از دیگری می‌شویم. (۱۰ نمره)
