



دانشگاه تهران

بسمه تعالی  
پایان ترم درس یادگیری ماشین  
بهمن ماه ۱۴۰۲

نام و نام خانوادگی: ..... شماره دانشجویی: .....



پردیس دانشکده‌های فنی

۱. خوشه‌یابی به روش k-means را در نظر بگیرید:

الف) شرط انتقال یک نمونه از یک خوشه به خوشه دیگر را به دست آورید. (منظور استخراج شرط با ذکر مراحل ریاضی آن است) (۵ امتیاز).

ب) یک الگوریتم یا فلوچارت (کامل و واضح) برای خوشه‌یابی به روش k-means ارائه دهید (۵ امتیاز).

ج) آیا کمینه کردن تابع هزینه در خوشه‌یابی به روش k-means صرفاً فاصله درون کلاسی را کم می‌کند یا فاصله بین کلاسی را هم زیاد می‌کند؟ استدلال کنید (۵ امتیاز).

۲. چارچوب کلی LDA و PCA را در ذهن داشته باشید:

الف) ابتدا نشان دهید (منظور اثبات رابطه با ذکر مراحل ریاضی آن است) (۱۰ امتیاز)

$$S_T = S_W + S_B$$

ب) با توجه به این رابطه استدلال کنید که چرا برای محاسبه ماتریس جداپذیری ابتدا باید الگوریتم PCA را اجرا کرده و در صورت لزوم کاهش بعد بدون سرپرستی را انجام دهیم (۵ امتیاز).

ج) آیا امکان دارد بعد از مرحله (ب) هنوز  $S_W$  وارون پذیر نباشد؟ استدلال کنید (۵ امتیاز).

۳. روش انتخاب پیش‌رونده (Forward Selection) برای انتخاب ویژگی را در نظر بگیرید:

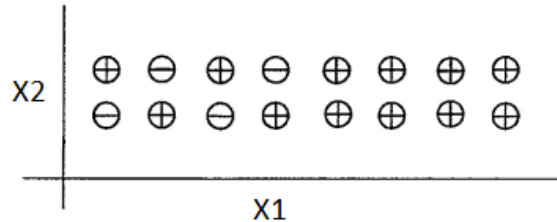
الف) یک الگوریتم یا فلوچارت (کامل و واضح) برای روش انتخاب پیش‌رونده ارائه دهید. شرط توقف را شفاف بیان کنید (۱۰ امتیاز).

ب) آیا ممکن است اولین ویژگی انتخاب شده در این روش پس از چند گام دیگر مطلوب‌ترین ویژگی نباشد؟ استدلال کنید (۵ امتیاز).

ج) یک راه حل برای کاهش مشکل مطرح شده در بند (ب) ارائه دهید (۵ امتیاز).

د) آیا حذف پس‌رونده (Backward Elimination) هم دچار مشکل مشابه مطرح شده در بند (ب) هست؟ استدلال کنید (۵ امتیاز).

۴. مجموعه داده‌های زیر را که شامل ۱۲ نمونه مثبت و ۴ نمونه منفی است در نظر بگیرید. روی این مجموعه داده دو درخت تصمیم آموزش می‌دهیم. درخت اول که نامش **underfit** است، فقط شامل یک رأس برگ است (کل درخت شامل یک رأس است). درخت دوم که نامش **overfit** است، به صورت استاندارد روی این مجموعه داده ساخته می‌شود و هیچ‌گونه الگوریتم هرسی روی آن اعمال نمی‌شود.



- الف) درخت **overfit** دقیقاً چند راس برگ و چند رأس داخلی دارد؟ توضیح دهید (۵ امتیاز).
- ب) روش **leave-one-out** برای درخت **overfit** روی این مجموعه داده چقدر خطا خواهد داشت؟ در واقع توضیح دهید که تعداد **misclassification** ها در این حالت چند مورد است (۵ امتیاز).
- ج) به سوال بند (ب) برای درخت **underfit** جواب دهید (۵ امتیاز).

۵. فرض کنید مجموعه داده  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  را در اختیار داریم. می‌خواهیم یک رگرسیون خطی با استفاده از نسخه ساده شده‌ی **Support vector regression** به صورت زیر انجام دهیم:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{subject to: } \begin{cases} y_i - w^T x_i - b \leq \epsilon & i = 1, \dots, n \\ w^T x_i + b - y_i \leq \epsilon & i = 1, \dots, n \end{cases}$$

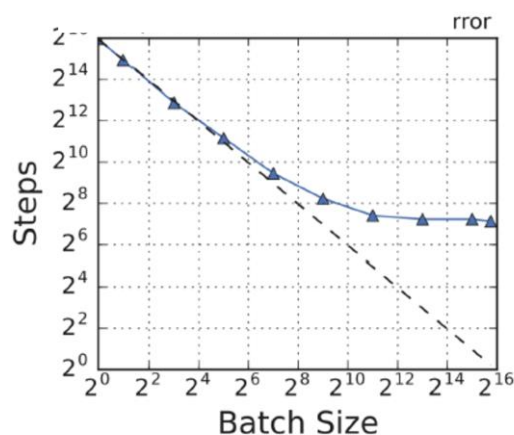
شهود رابطه‌ی بهینه‌سازی بالا این است که می‌خواهیم یک مسأله رگرسیون خطی به صورت  $y = w^T x + b$  را حل کنیم (یعنی  $w$  و  $b$  را بدست آوریم)، به طوری که قدر مطلق فاصله‌ی نقاط تا خط بدست آمده کمتر از مقدار  $\epsilon$  باشد و از طرفی نرم ۲ بردار  $w$  بزرگ نباشد تا دچار **overfitting** نشویم.

الف) تابع لاگرانژین مسأله بهینه‌سازی بالا را تشکیل دهید (۱۰ امتیاز).

ب) مسأله دوگان را برای مسأله بهینه‌سازی بالا بدست آورید. (نیازی به حل مسأله دوگان نیست فقط صورت آن را بدست بیاورید) (۲۰ نمره)

۶. مجموعه داده‌هایی با برچسب باینری در نظر بگیرید که به کمک یک طبقه بند خطی امکان جداپذیری ندارد. یک شبکه عصبی را در نظر بگیرید که از یک لایه پیمشی (Convolution layer) با یک تابع فعال سازی خطی تشکیل شده است و در انتها یک لایه خطی با خروجی لجستیک (Logistic) وجود دارد. آیا به کمک چنین معماری می‌توان تمام نمونه‌های آموزشی را به طور کامل طبقه بندی کرد؟ پاسخ را توضیح دهید؟

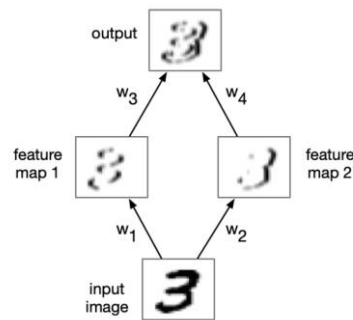
۷. نمودار زیر نشان دهنده تعداد تکرارهای مورد نیاز در روش گرادیان نزولی تصادفی ( Stochastic Gradient Descent) برای رسیدن به هزینه (Loss) معین، بر حسب تابعی از اندازه دسته است.



الف) برای اندازه‌های دسته کوچک، تعداد تکرارهای مورد نیاز برای رسیدن به هزینه معین با افزایش اندازه دسته، کاهش می‌یابد. چرایی این پدیده را توضیح دهید (۵ امتیاز).

ب) برای اندازه‌های بزرگ، با افزایش اندازه دسته، تعداد تکرارها تغییر چندانی نمی‌کند. چرا اینطور است (۵ امتیاز)؟

۸. معماری شبکه کانولوشنال زیر را برای تشخیص مرزهای عمودی در یک تصویر در نظر بگیرید.



تابع فعال سازی ReLU در اولین لایه پیچشی اعمال می شود. لایه خروجی از تابع فعال سازی خطی استفاده می کند. در تصاویر از سفید برای نشان دادن ۰ و مقادیر تیره تر برای نشان دادن مقادیر بزرگتر (مثبت تر) استفاده می کنیم. الف) دو فیلتر کانولوشن برای لایه اول، به اندازه  $3 \times 3$  طراحی کنید. یکی از آنها باید مرزهای تاریک/روشن را تشخیص دهد و دیگری باید مرزهای روشن/تاریک را تشخیص دهد. پاسخ خود را توضیح دهید (۱۰ امتیاز).

ب) یک فیلتر کانولوشن را با اندازه  $3 \times 3$  برای لایه خروجی طراحی کنید، که خروجی مورد نظر را محاسبه می کند. نیازی به توضیح پاسخ نیست (۵ امتیاز)..

۹. الف) شباهت روش **bagging** و **bootstrap** را با ذکر یک مثال به صورت خلاصه توضیح دهید (۵ امتیاز)..

ب) اثر همبستگی بین نمونه‌ها بر روی واریانس مدل در روش **bagging** چگونه است؟ پاسخ خود را به صورت خلاصه شرح دهید (۵ امتیاز).