



یادگیری ماشین

پاییز ۱۴۰۳

استاد: علی شریفی زارچی

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

مدت آزمون: ۱۸۰ دقیقه

نمونه سوال میانترم اول

تاریخ آزمون: ۲۴ آذر

۱. (۲۵ نمره)

فرض کنید مجموعه داده آموزشی $\{(X_i, Y_i)\}_{i=1}^n$ داریم که در آن $X_i \in \mathbb{R}^p$ و $Y_i \in \{0, 1\}$ هستند. چگالی شرطی X با فرض $Y = k$ را با $g_k(x)$ نشان می‌دهیم و فرض می‌کنیم که برای هر $x \in \mathbb{R}^p$ ، $g_k(x) > 0$ است. همچنین احتمال هر کلاس را با $\pi_k = P(Y = k)$ تعریف می‌کنیم و $q(x) = P(Y = 1 | X = x)$ را معرفی می‌کنیم.

(الف) دسته‌بند بیز که ریسک را با استفاده از زیان صفر و یکی $\mathbb{I}(f(X) \neq Y)$ کمینه می‌کند، به صورت زیر تعریف می‌شود:

$$f_{\text{Bayes}}(x) = \arg \max_{k \in \{0, 1\}} \pi_k g_k(x)$$

زیان مورد انتظار شرطی $\mathbb{P}(f(X) \neq Y | X = x)$ را برای یک داده تست $X = x$ بر حسب $q(x)$ بنویسید. [نتیجه باید فقط به $q(x)$ وابسته باشد.]

(ب) با داشتن یک نقطه تست $X = x$ و نزدیک‌ترین همسایه آن $X' = x'$ ، زیان مورد انتظار شرطی $\mathbb{P}(f_{\text{NN1}}(X) \neq Y | X = x, X' = x')$ را برای دسته‌بند NN۱- بر حسب $q(x)$ و $q(x')$ محاسبه کنید. (در NN۱- ما برچسب نزدیک‌ترین همسایه در مجموعه داده آموزشی را به عنوان برچسب اعلام می‌کنیم.)

(ج) با افزایش تعداد نمونه‌های آموزشی به بی‌نهایت، یعنی $n \rightarrow \infty$ ، فرض کنید که داده‌های آموزشی فضا را به گونه‌ای پر می‌کنند که $q(x') \rightarrow q(x)$ برای هر x . حد $n \rightarrow \infty$ از $\mathbb{P}(f_{\text{NN1}}(X) \neq Y | X = x)$ را پیدا کنید. اگر $R_{\text{Bayes}} = \mathbb{P}(Y \neq f_{\text{Bayes}}(X))$ و $R_{\text{NN1}} = \mathbb{P}(Y \neq f_{\text{NN1}}(X))$ را تعریف کنیم، نشان دهید که برای n کافی بزرگ، داریم:

$$R_{\text{Bayes}} \leq R_{\text{NN1}} \leq 2R_{\text{Bayes}}(1 - R_{\text{Bayes}})$$

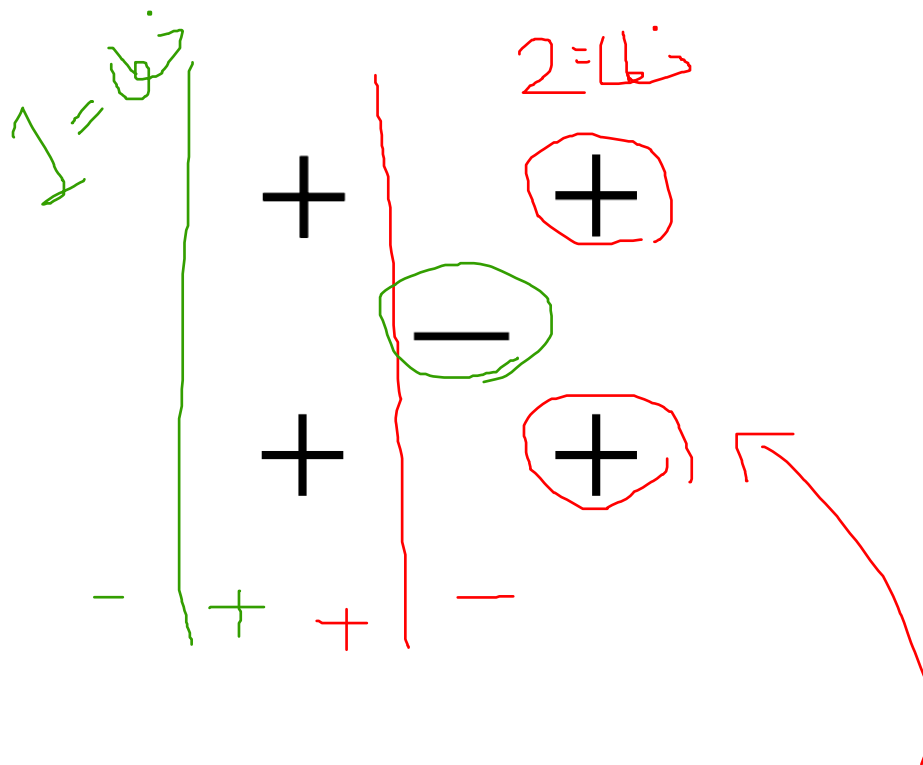
۲. (۲۵ نمره) یک مدل $\text{regression logistic}$ در نظر بگیرید. می‌دانیم $y = g(w^T x)$ و تابع هزینه آن $\text{binary entropy cross}$ است و

$$g(z) = \frac{1}{1 + e^{-z}} \quad \text{و} \quad g'(z) = g(z)(1 - g(z))$$

یک روش تغییر داده شده از $\text{regression logistic}$ ارائه می‌کنیم که در آن $g(z) = \frac{e^{-z}}{1 + e^{-z}}$ و تابع هزینه همچنان همان $\text{binary entropy cross}$ است. پارامترها و پیش‌بینی‌های مدل یاد گرفته شده جدید نسبت به مدل اولیه چه تفاوتی دارد؟ به صورت ریاضی توضیح دهید.

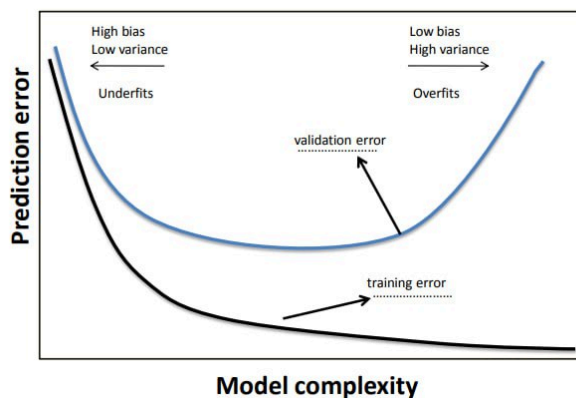
$$y' = 1 - y$$

۳. (۲۵ نمره) فرایند آموزش یک دسته‌بند boosting با استفاده از decision stumps ها را روی مجموعه داده زیر در نظر بگیرید:



- وزن کدام یک از داده ها بعد از اتمام اولین *iteration* افزایش می یابد ؟ دور آن ها را خط بکشید.
- چند *iteration* طول می کشد تا روی داده آموزش، خطایی برابر با صفر داشته باشیم ؟ *3 تا 4*
- آیا می توانید یک نمونه دیگر روی داده ها اضافه کنید تا الگوریتم *boosting* بعد از دو گام به خطای آموزش صفر برسد ؟ در صورت ممکن نبودن دلایل خود را توضیح دهید.
- چرا در الگوریتم *boosting* ما از *weaklearner* ها استفاده می کنیم ؟

۴. (۲۵ نمره) شکل زیر نشان دهنده منحنی های *training* و *validation* یک مدل بر اساس افزایش پیچیدگی آن را نشان می دهد. با توجه به این شکل، به سوالات زیر پاسخ دهید.



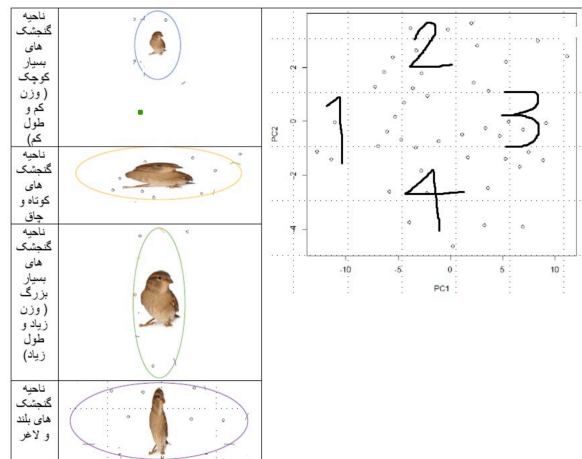
- کدام یک از منحنی ها شبیه *trainigerror* و کدام یک *validaitonerror* هستند ؟
- در کدام نواحی این شکل بایاس و واریانس کم یا زیاد هستند ؟ ۴ قسمت نشان دهید که هر کدام نشان دهنده واریانس کم، واریانس بالا، بایاس کم و بایاس زیاد باشد.
- در کدام نواحی این مدل *over fit* یا *under fit* می شود ؟

۵. (۲۵ نمره) اگر X_1 و X_2 به ترتیب طول و وزن تعدادی گنجشک باشند که اندازه های آن ها ثبت شده است و ماتریس واریانس-کوواریانس آن ها به صورت زیر محاسبه شده است :

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

وزن طول
طول وزن

ابتدا مولفه اصلی اول PC_1 و مولفه اصلی دوم PC_2 را که به صورت ترکیب خطی X_1 و X_2 هستند را بر اساس مقادیر ویژه و بردار ویژه نرمال شده این ماتریس به دست آورید. هر کدام چه درصدی از اطلاعات موجود در متغیرهای اصلی را بیان می کنند؟ همچنین هر کدام از گنجشک ها با شرایط گفته شده در شکل، در کدام ناحیه (نواحی ۱ تا ۴) مختصات دو بعدی (PC_1, PC_2) قرار می گیرند؟



۶. (۲۵ نمره)

با توجه به الگوریتم Adaboost به سوالات زیر پاسخ دهید.

(آ) AdaBoost را می توان به عنوان یک بهینه ساز در نظر گرفت که تابع خطای نمایی $E = \sum_{i=1}^N \exp(-y_i f(x_i))$ را کمینه می کند؛ جایی که $y = +1$ یا $y = -1$ برچسب کلاس، x داده و $f(x)$ جمع وزنی از weak learnerهاست. نشان دهید که تابع خطا E همواره از تابع خطای $0 - 1$ یعنی

$$E_{0-1} = \sum_{i=1}^N 1 \cdot (y_i f(x_i) < 0)$$

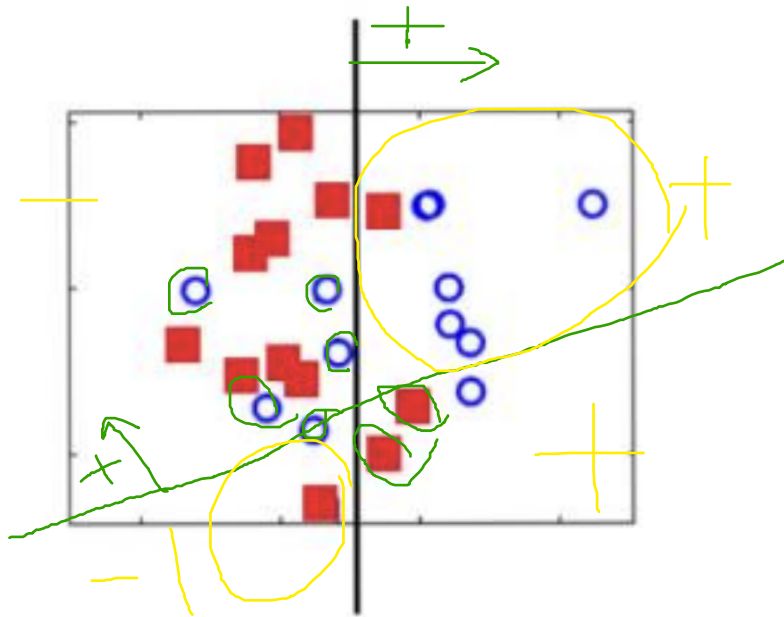
بزرگتر است و در نتیجه یک کران بالا برای آن محسوب می شود (راهنما: E_{0-1} تابعی است که برای پیش بینی درست مقدار ۱ و در غیر این صورت مقدار ۰ را برمی گرداند).

(ب) الگوریتم AdaBoost دو محدودیت دارد. با در نظر گرفتن این دو محدودیت به سوالات زیر پاسخ دهید.

• به صورت ریاضی توضیح دهید که چرا یک weak learner با دقت کمتر از ۵۰٪ برای AdaBoost مشکل ساز است.

• حساسیت AdaBoost به outlierها را در نظر بگیرید. یک راهکار ساده برای کاهش این مشکل پیشنهاد کنید.

(ج) شکل زیر مرز تصمیم‌گیری (خط میانی) پس از اولین تکرار در یک طبقه‌بند AdaBoost با decision stumpها به عنوان weak learner را نشان می‌دهد. نقاط مربعی از کلاس ۱- و دایره‌ها از کلاس ۱+ هستند. مرز تصمیم‌گیری در تکرار دوم را تقریبی با یک خط پیوسته ترسیم کنید و مرز تصمیم‌گیری نهایی را بر اساس تصمیمات در تکرارهای ۱ و ۲ با خط چین نشان دهید. توضیح خود را ارائه دهید.



شکل ۲: خط میانی مرز تصمیم‌گیری پس از اولین تکرار در AdaBoost را نشان می‌دهد. طبقه‌بند نقاط سمت چپ خط را به عنوان کلاس ۱- و نقاط سمت راست را به عنوان کلاس ۱+ پیش‌بینی می‌کند.