



یادگیری ماشین

بهار ۱۴۰۳

استاد: فاطمه سیدصالحی

تمرین چهارم

مهلت ارسال: ۹ خرداد

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- همکاری و همفکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت همفکری و یا استفاده از هر منابع خارج درسی، نام همفکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- موضوعات تمرین: درخت تصمیم و یادگیری مبتنی بر نمونه و یادگیری جمعی

سوالات نظری (۶۵ نمره)

۱. (۱۰ نمره) فرض کنید می‌خواهیم درخت تصمیمی بسازیم که پیش‌بینی کند در چه بازی‌هایی هوش مصنوعی می‌تواند به سطحی برسد که بهتر از انسان بازی کند. جدول زیر دیتاست شما را نمایش می‌دهد که در آن برای هر نمونه ۳ ویژگی x_1 ، x_2 و x_3 را داریم. همچنین خروجی دیتاست را با y نمایش می‌دهیم.

x_1 (Team or Individual)	x_2 (Mental or Physical)	x_3 (Skill or Chance)	y (Win or Lose)
T	M	S	W
I	M	S	W
T	P	S	W
I	M	C	W
T	P	S	L
I	M	C	L
T	P	C	L
T	P	C	L
T	P	C	L
I	P	S	W

الف) آنتروپی کلاس y ($H(y)$) را محاسبه کنید.

ب) بر حسب معیار information gain بهترین درخت با عمق ۱ را بسازید و محاسبات را به صورت کامل نشان دهید. بر اساس کدام ویژگی split در ریشه درخت انجام می‌شود؟

۲. (۲۰ نمره) الگوریتم AdaBoost را در نظر بگیرید. در این الگوریتم T دسته بند با نام‌های G_1, \dots, G_T به صورت تکرار شونده آموزش داده می‌شوند و در هر مرحله وزن نسبت داده شده به داده‌های آموزش تغییر می‌کند. فرض کنید z یک نمونه داده تست باشد. خروجی metalearner (مدل اصلی) ترکیب خطی از خروجی دسته‌بندها می‌باشد:

$$M_T(z) = \sum_{t=1}^T \beta_t G_t(z)$$

که $M_T(z)$ خروجی metalearner و $G_t(z)$ خروجی دسته بند t ام روی نمونه z است.

مدل AdaBoost، metalearner را با تابع هزینه نمایی آموزش می‌دهد. در این سوال الگوریتم boosting دیگری به نام Gradient Boosting را بررسی می‌کنیم. فرض کنید $\rho = M_T(z)$ خروجی metalearner در این الگوریتم و $\ell \in \{+1, -1\}$ لیبیل واقعی برای نمونه z باشند. الگوریتم Gradient Boosting تابع هزینه نمایی را با squared error loss جایگزین می‌کند:

$$L(\rho, \ell) = (\rho - \ell)^2$$

فرض کنید که دسته بندی با ۲ کلاس انجام می‌شود و n داده آموزشی $X_1, X_2, \dots, X_n \in \mathbb{R}^d$ و بردار $y \in \mathbb{R}^n$ از برچسب‌ها را داریم که $y_i = \pm 1$ است.

الف) عبارتی برای تابع هزینه این مدل بنویسید. دقت کنید که براساس توضیحات داده شده تابع هزینه از جنس MSE (mean squared error) است. پاسخ خود را بر حسب β_t ها، G_t ها و y_i ها بیان کنید.

ب) فرض کنید در مرحله T ام از الگوریتم Gradient Boosting هستیم و دسته بند ضعیف G_T را آموزش داده‌ایم ولی هنوز ضریب β_T را تعیین نکرده‌ایم. همچنین فرض کنید ضرایب $\beta_1, \dots, \beta_{T-1}$ در مرحله T ام ثابت مانده و دچار تغییر نمی‌شوند. مقدار β_T را طوری تعیین کنید که تابع هزینه metalearner را کمینه کند. عبارت شما باید تا حد ممکن ساده شده باشد.

۳. (۱۰ نمره) به سوالات زیر پاسخ دهید.

الف) اگر از bootstrapping استفاده کنیم تا $N' = pN$ نمونه از N داده نمونه‌گیری کنیم، ثابت کنید به طور تقریبی $e^{-p} \cdot N$ از داده‌ها به طور کلی در نمونه‌گیری انتخاب نخواهند شد. (فرض کنید N بسیار بزرگ است.)

ب) یک جنگل تصادفی مانند G در نظر بگیرید به طوریکه از سه درخت تصمیم باینری $\{g_k\}_{k=1}^3$ تشکیل شده است. فرض کنید هر کدام از این درخت‌های تصمیم بر روی داده تست دارای خطای زیر می‌باشند:

$$E_{out}(g_1) = 0.15, E_{out}(g_2) = 0.25, E_{out}(g_3) = 0.35$$

محدوده‌ی خطای $E_{out}(G)$ را با ذکر دلیل بدست آورید.

۴. (۸ نمره) کدام یک از موارد زیر درباره‌ی bagging, boosting درست می‌باشد؟ (می‌تواند چند مورد درست باشد)

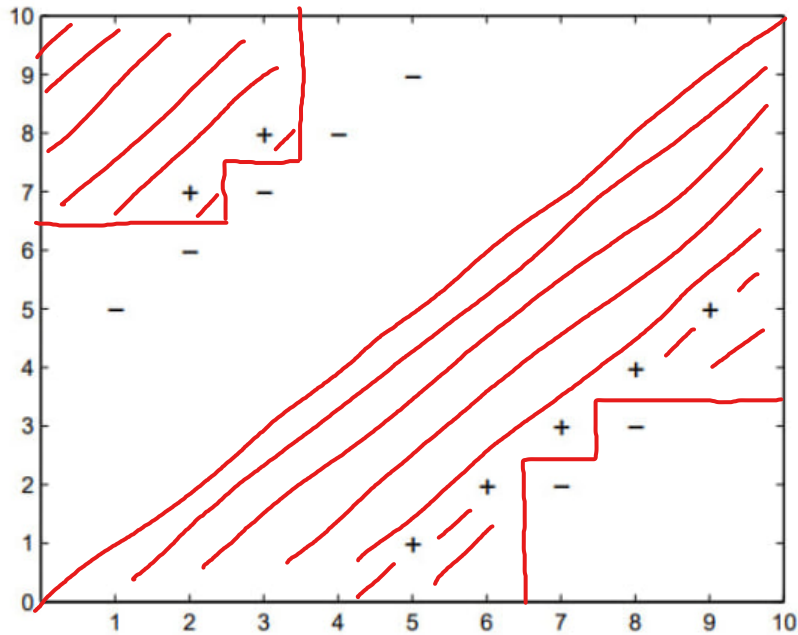
الف) یادگیرنده‌های مختلف در bagging می‌توانند به صورت موازی آموزش داده شوند.

ب) یادگیرنده‌های مختلف در boosting می‌توانند به صورت موازی آموزش داده شوند.

ج) هر یک از دسته‌بندها در bagging بر روی کل داده آموزش، آموزش داده می‌شوند.

د) هر یک از دسته‌بندها در boosting بر روی کل داده آموزش، آموزش داده می‌شوند.

۵. (۱۲ نمره) در این سوال یک دسته بند KNN با متریک فاصله L_2 را در نظر بگیرید. کلاس‌ها را تماماً باینری در نظر خواهیم گرفت. به سوالات زیر با توجه به دیتاست مشخص شده پاسخ دهید.



- الف) به ازای چه مقدار K خطای این دسته بند کمینه می شود؟ مقدار این خطا چقدر است؟
- ب) چرا استفاده از مقادیر خیلی زیاد یا خیلی کم برای K روی این دیتاست می تواند اشتباه باشد؟
- ج) فرض کنید از روش leave one out cross validation استفاده می کنیم. به ازای چه مقدار از K خطا کمینه می شود؟ مقدار این خطا چقدر است؟
- د) مرز تصمیم دسته بند 1NN classifier را برای این دیتاست در تصویر نمایش دهید.

۶. (۵ نمره) فرض کنید T مجموعه ای از n داده آموزشی با ویژگی های A_1, \dots, A_a و k کلاس c_1 تا c_k باشد. حال یک ویژگی خاص به نام A را در نظر بگیرید. فرض کنید مجموعه T را بر اساس این ویژگی split کرده و یک جداسازی disjoint و کامل روی T می سازیم. این جداسازی به صورت $\{T_i^A | i \in \{1, \dots, m_A\}\}$ خواهد بود که m_A تعداد حالات مختلف برای ویژگی A است. همچنین فرض کنید توزیع کلاس های T به صورت یکنواخت و مستقل از ویژگی A باشد. موارد زیر را بر حسب پارامتر k (تعداد کلاس ها) محاسبه کنید و پاسخ خود را تفسیر کنید.

$$H(T), H(T|A), IG(T, A)$$

سوالات عملی (۳۵ نمره)

۱. (۳۵ نمره) برای حل سوالات به notebook های ضمیمه شده مراجعه کنید.