



دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

بهار ۱۴۰۰

تمرین سری سوم

مدرس: دکتر محمدحسین رهبان

زمان تحویل: ۲۰ فروردین

سوال ۱ Bias-Variance Tradeoff (۲۵ نمره)

آ) تجزیه‌ی *Bias-Variance* را به زبان ریاضی و به صورت کامل بنویسید. (۵ نمره)

ب) با در نظر گرفتن محورهای خطا^۱ و پیچیدگی مدل^۲، نمودارهای اریبی^۳، واریانس و خطای کل^۴ را رسم کنید. (۲ نمره)

پ) مدلی با مقدار زیادی اریبی را در نظر بگیرید. دست‌کم دو راهکار برای رفع و یا کم کردن اثر اریبی پیشنهاد دهید. (۵ نمره)

ت) همان‌طور که می‌دانید تجزیه‌ی *Bias-Variance* یک تحلیل کلی‌ست و بنا به ویژگی‌های منحصر به فرد هر مجموعه‌داده، ممکن است با مشکلاتی که از مقدار زیاد اریبی یا واریانس زیاد پیش‌بینی می‌کنیم، مواجه نشویم. فرض کنید تعداد ویژگی‌های یک مجموعه‌داده را به ندرت کم می‌کنیم (به عنوان مثال ویژگی‌هایی که حالت نویز دارند). با تحلیل از دیدگاه *Bias-Variance* بگویید هر یک چه تغییری می‌کنند؟ (۵ نمره)

ث) همکار شما در آزمایشگاه، مدلی را در نظر گرفته که واریانس زیادی دارد و پس از پیاده‌سازی و مشاهده‌ی نتیجه‌ها متوجه می‌شوید که این مقدار بیش از حد زیاد است. دست‌کم دو راهکار برای رفع اثر منفی این مقدار واریانس پیشنهاد دهید. (۵ نمره)

ج) اگر تعدادی از ویژگی‌ها دوه‌دو با یکدیگر هم‌بسته باشند، از دیدگاه *Bias-Variance* این اتفاق چطور تعبیر می‌شود؟ حذف ویژگی‌های هم‌بسته اریبی را تغییر می‌دهد یا واریانس را؟ آن را افزایش می‌دهد یا کاهش؟ (۳ نمره)

راهنمایی: به صورت شهودی، اریبی به معنی جهت‌دهی شده و جلوگیری از بروز ارتباط ویژگی‌ها و مقدار خروجی است و واریانس به معنی میزان عملکرد مناسب روی داده‌هایی‌ست که تا به حال دیده نشده‌اند.

سوال ۲ Linear Regression (۱۵ نمره)

در رویکردی برای مسئله‌ی رگرسیون خطی، تلاش می‌کنیم بهترین پارامتر مانند w_1 را تخمین بزنیم (با استفاده از آموزش روی

¹Error

²Model Complexity

³Bias

⁴Total Error

مجموعه داده‌هایمان) و سپس با بکار گرفتن این پارامتر پا به دنیای واقعی گذاشته و پیش‌بینی‌هایمان را انجام می‌دهیم. حال یکی از دوستان شما بنا به مشاهداتش، می‌گوید دومین بهترین پارامتر به نام w_2 به دلیل درصد بسیار کم و ناچیزی دوم شده‌است و تا حد بسیار خوبی نسبت به w_1 عملکرد مناسبی دارد. w_3 هم چنین وضعی نسبت به w_2 دارد و الی آخر. او با توجه به مشاهداتش ادعا می‌کند بجای در نظر گرفتن تنها یک پارامتر (بهترین پارامتر)، تمام پارامترها را متناسب با عملکردشان در نظر بگیریم (از نظر وی w یک متغیر تصادفی با توزیع گاوسی است). سپس با دیدن مجموعه داده‌هایمان، توزیع این متغیر تصادفی را به روز کرده و از آن برای پیش‌بینی‌ها استفاده کنیم. آیا این رویکرد از دیدگاه *Bias-Variance* نسبت به رویکرد فعلی‌تان بهتر است؟ چرا؟

• وضعیت فعلی شما:

$$w^* = \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \mathbb{P}(w|data)$$

که ممکن است آن را به روش *ML* و یا *MAP* محاسبه کنید.

• وضعیت پیشنهادی دوست‌تان:

$$\mathbb{P}(w|data) = \frac{\mathbb{P}(w) \times \mathbb{P}(data|w)}{\mathbb{P}(data)}$$

سوال ۳ *VC Dimension* (۲۳ نمره)

(آ) بعد *VC* را برای هریک از دسته‌بندی‌های زیر بدست آورید: (۱۵ نمره)

۱. H_r : مجموعه همه مستطیل‌هایی که ضلع‌هایشان موازی محورهای مختصات صفحه \mathbb{R}^2 است. نمونه‌های مثبت درون مستطیل قرار می‌گیرند.

۲. H_c : مجموعه همه دایره‌هایی که در صفحه \mathbb{R}^2 وجود دارند. نمونه‌های مثبت درون دایره قرار می‌گیرند.

۳. مجموعه همه ابرصفحه‌های گذرنده از مبدا^۱ در فضای \mathbb{R}^d .

(ب) چند نمونه آموزش کافی است تا اطمینان حاصل کنیم که می‌توان با احتمال ۰.۹ یک یادگیرنده^۲ از فضای فرضیه H_c را با کمینه دقت ۰.۹۵ آموزش داد؟ (۴ نمره)

(پ) وقتی می‌گوییم یادگیرنده با احتمال ۰.۹ موفق خواهد شد، منظور از قسمت ب چیست؟ برای پاسخ‌گویی به این سوال توصیفی از یک آزمایش ساده را ارائه دهید که می‌توان آن را به دفعات تکرار کرد و احتمال موفقیت در آن دست کم ۰.۹ است. (۴ نمره)

¹Origin Crossing Hyperplanes

²Learner

سوال ۴ سؤال عملی - Kernel Regression (۲۷ نمره)

در یکی از رویکردهای معمول برای مسئله‌ی رگرسیون، با استفاده از داده‌ی آموزش، مدل را آموزش داده و پس از آن تنها با پارامترهایی که در اختیار داریم پا به دنیای واقعی گذاشته و پیش‌بینی‌ها را انجام می‌دهیم (داده‌های آموزش را کنار می‌گذاریم). ولی رویکردهای مبتنی بر هسته^۱ حتی در زمان پیش‌بینی هم از داده‌های آموزش استفاده می‌کند. در این سوال می‌خواهیم با استفاده از تابع‌های هسته، مسئله‌ی رگرسیون را حل کرده و نقاط ضعف و قوت آن را بررسی نماییم. به این منظور از مجموعه داده‌ای که در اختیار شما قرار گرفته و مشخصات آن در فایل `data_description.txt` بیان شده، استفاده کنید.

آ) یکی از گام‌های مهم در حل مسائل با این رویکرد، انتخاب تابع هسته مناسب است. به عنوان گزینه‌ی اول تابع زیر را در نظر بگیرید:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{1}_{\|\mathbf{x}_i - \mathbf{x}_j\| \leq h}$$

با استفاده از تابع هسته به هر نقطه وزن خاصی اختصاص یافته و با توجه به مقدار y آن‌ها، پیش‌بینی انجام می‌شود. در این بخش رابطه وزن‌ها را بدست آورده و پس از پیاده‌سازی روش، مقدار میانگین مربعات خطا^۲ را گزارش دهید. در این تابع پارامتر h نقشی کلیدی دارد. به ازای h های مختلف پیش‌بینی را انجام داده و با رسم نمودار اثر افزایش و کاهش مقدار h بر میانگین مربعات خطا را بررسی کنید. (۷ نمره)

ب) حال به عنوان گزینه‌ی دوم تابع زیر را در نظر بگیرید:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma}}$$

۱. در این تابع، پارامتر σ نقشی تعیین‌کننده دارد. با در نظر گرفتن این تابع و پارامتر آن به بخش آ پاسخ دهید. (۷ نمره)

۲. آیا تعداد ویژگی‌ها بر عملکرد و یا انتخاب پارامتر σ تاثیری دارد؟ (۳ نمره)

۳. مقدار اریبی و واریانس را بر حسب پارامتر این مدل بدست آورید. (۲ نمره)

پ) نتیجه بخش آ و ب را با هم مقایسه کرده و دلیل برتری یا ضعف هر تابع هسته را بیان کنید. (۴ نمره)

ت) فرض کنید داده‌هایمان بسیار تنک^۳ هستند (مانند زمانی که بیشتر ستون‌ها *Categorical* هستند). به نظر شما روش رگرسیون خطی نتیجه‌ی بهتری خواهد داشت یا رگرسیون با هسته گاوسی؟ چرا؟ (۴ نمره)

پاینده باشید

^۱ Kernel

^۲ Mean Squared Error

^۳ Sparse