



یادگیری ماشین

بهار ۱۴۰۳

استاد: فاطمه سیدصالحی

مهلت ارسال: ۱۷ فروردین

تمرین دوم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ‌های هر کس حتماً 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.
- لطفاً تصویری واضح از پاسخ‌های سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- موضوعات تمرین: رگرسیون چند جمله‌ای، رگولاریزیشن، کرنل‌ها

سوالات نظری (۷۰ نمره)

۱. (۲۰ نمره) کرنل معتبر

با توجه به اطلاعات زیر مشخص کنید که هر یک از بخش‌ها می‌تواند یک کرنل باشند یا خیر، در صورت کرنل نبودن یک مثال نقض و در صورت کرنل بودن آنرا اثبات کنید.

$$a \in \mathbb{R}^+$$

$$f: \mathbb{R}^d \rightarrow \mathbb{R}^p$$

$$K_1, K_2 \text{ are kernels over } \mathbb{R}^d \times \mathbb{R}^d$$

$$K_2 \text{ is a kernel over } \mathbb{R}^p \times \mathbb{R}^p$$

$$K(x, z) = K_1(x, z) + K_2(x, z) \quad (\text{آ})$$

$$K(x, z) = aK_1(x, z) \quad (\text{ب})$$

$$K(x, z) = K_2(f(x), f(z)) \quad (\text{ج})$$

۲. (۳۰ نمره) رگولاریزیشن

در نظر بگیرید که می‌خواهیم مسئله دسته‌بندی باینری نشان داده شده در شکل ۱ را با استفاده از یک مدل ساده رگرسیون لجستیک حل کنیم.

$$P(y = 1 | \mathbf{x}, \mathbf{w}) = g(w_0 + w_1 x_1 + w_2 x_2) = \frac{1}{1 + \exp(-w_0 - w_1 x_1 - w_2 x_2)}$$

همانطور که از روی شکل هم مشخص است، داده‌های آموزشی را می‌توان با صفر خطای آموزشی از یکدیگر جدا کرد. حال می‌خواهیم برای حل مسئله دسته‌بندی مقدار عبارت زیر را برای مقادیر بزرگ C حداکثر کنیم.

$$\sum_{i=1}^n \log P(y_i | x_i; w_0, w_1, w_2) - C w_2^2$$

در این عبارت $C w_2^2$ ، بخش تعدیل‌گر^۱ است که $j \in \{0, 1, 2\}$ به عبارتی تنها یکی از پارامترها تعدیل می‌شوند. با توجه به داده‌های آموزشی داده شده در شکل یک بیان کنید که در هریک از حالات زیر خطای آموزشی چه تفاوتی با مدل رگرسیون لجستیک ساده خواهد داشت؟ برای جواب خود استدلال بیاورید.

^۱ Regularization term

آ) با تعدیل w_2

ب) با تعدیل w_1

ج) با تعدیل w_0

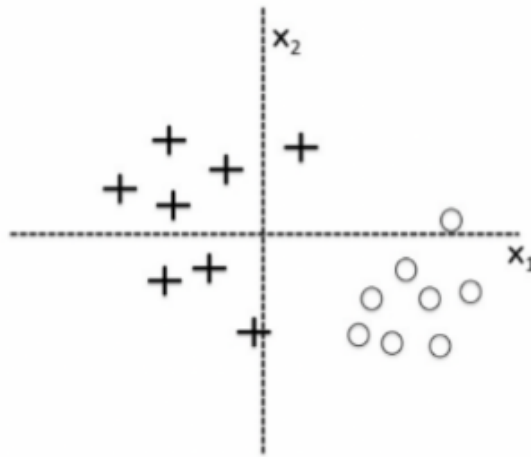
حالا در نظر بگیرید که می‌خواهیم هر دو پارامتر w_1 و w_2 را تعدیل کنیم. به عبارتی در مدل خود می‌خواهیم عبارت زیر را حداکثر کنیم

$$\sum_{i=1}^n \log P(y_i | x_i; w_0, w_1, w_2) - C(w_1^2 + w_2^2)$$

داده‌های آموزشی نیز همان داده‌های شکل ۱ هستند.

د) برای مقادیر بزرگ C ، انتظار داریم w_0 چه مقادیری بگیرد؟ استدلال بیاورید.

ه) اینبار در نظر بگیرید که تعدادی داده "+" که در کلاس $y = 1$ قرار دارند را به داده‌های آموزشی خود اضافه می‌کنیم. با فرض اینکه همچنان دو کلاس به طور کامل قابل جداسازی باشند، استدلال کنید که چه مقادیری برای w_0 انتظار داریم.



شکل ۱: داده‌های + متعلق به کلاس $y = 1$ و داده‌های O متعلق به کلاس $y = 0$ هستند.

۳. (۲۰ نمره) کرنل گوسی، کرنل معتبر

الف) نشان دهید که کرنل Gaussian را می‌توان به صورت ضرب داخلی بردارهای ویژگی با بعد نامتناهی نوشت:

$$k(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}}$$

ب) ماتریس $A \in R^{p \times p}$ را یک ماتریس متقارن و مثبت نیمه معین در نظر بگیرید. ثابت کنید

$$k(x, y) = x^T A y$$

یک تابع کرنل معتبر است.

۱. (۲۰ نمره) در این سوال می‌خواهیم مسئله‌ی رگرسیون را برای چندجمله‌ای‌ها بررسی کنیم. برای این کار طبق نوتبوک مراحل زیر را طی خواهید کرد:

- با استفاده از معادله‌ی داده شده یک مجموعه داده همراه با نویز گاوسی آماده کنید.
- داده‌ها را به دو مجموعه‌ی آموزش و آزمون تقسیم کنید. همچنین یک مجموعه داده‌ی آزمون بسازید که خارج از محدوده‌ی مجموعه‌ی آموزش ما باشد.
- با استفاده از مجموعه‌ی آموزش مدل‌های رگرسیون چندجمله‌ای با درجات مختلف را آموزش دهید.
- با استفاده از خطای کمترین مربعات، عملکرد این مدل‌ها را برای هر سه مجموعه بسنجید و با نمودار نشان دهید.
- در بخش آخر، طبق نمودارها و عملکرد مدل‌ها به سوال پرسیده شده پاسخ دهید.

۲. (۲۰ نمره) در این سوال، شما با مجموعه داده‌ای کار خواهید کرد که شامل اطلاعات خانه‌های بوستون است که در سال ۱۹۷۸ جمع‌آوری شده است و هر یک از ۵۰۶ ورودی نشان‌دهنده داده‌های تجمعی درباره ۱۴ ویژگی برای خانه‌های مختلفی از شهر بوستون و حومه آن هستند. متغیر هدف اصلی در این مجموعه داده، مقدار میانی ارزش خانه‌های متعلق به مالکان مقیم (MEDV) است که به عنوان یک شاخص برای تعیین قیمت‌خانه‌ها عمل می‌کند. ابتدا، با استفاده از کتابخانه scikit-learn (sklearn) یک مدل رگرسیون ساده را ایجاد می‌کنید. از این مدل رگرسیون به عنوان نقطه مرجع استفاده خواهیم کرد.

در ادامه، شما از تکنیک‌های عادی سازی Ridge و Lasso برای بهبود پیش‌بینی‌های خود استفاده کرده و تأثیر این روش‌ها بر دقت نتایج خود را ارزیابی خواهید کرد. عادی سازی روشی است که برای جلوگیری از بیش‌برازش (overfitting) که با اضافه کردن یک جریمه به تابع Loss انجام می‌شود. روش Lasso مقادیر مطلق ضرایب را به عنوان جریمه اضافه می‌کند، در حالی که روش Ridge مربع ضرایب را به عنوان جریمه اضافه می‌کند. با ادغام این تکنیک‌های عادی سازی، شما این هدف را دارید که عملکرد تعمیمی مدل‌های رگرسیون خود را بهبود بخشیده و پیش‌بینی‌های بهتری از قیمت خانه در منطقه بوستون بدست آورید.