

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

درس یادگیری ماشین

تمرین سری اول

امیر خزاعی - ۹۶۱۳۱۰۳۸

پاییز ۱۳۹۶

سؤال اول

یادگیری بانظارت

این دسته از الگوریتم‌ها به کمک داده‌های برچسب خورده سعی در برچسب‌زنی داده‌های برچسب نخورده دارند. به طور مثال فرض کنید اطلاعات هزار بیمار سرطانی را در اختیار دارید. اندازه تومور، سن و ... پارامترهایی از اطلاعات هر بیمارند. درنهایت در بیمار برچسب «سرطان خوش‌خیم» و یا «سرطان بدخیم» را خواهد داشت. حال با استفاده از یادگیری بانظارت می‌توان از روی اطلاعات بیمار (X) و برچسب هر کدام (Y) تابع $f(x) = y$ را محاسبه نمود. حال می‌توان نوع سرطان بیماران جدید را به راحتی و با صرف هزینه پایین تشخیص داد.

الگوریتم‌های یادگیری بانظارت به دو بخش دسته‌بندی¹ و رگرسیون² تقسیم می‌شوند. در الگوریتم‌های دسته‌بندی سعی در پیش‌بینی مقادیر گسسته و در الگوریتم‌های رگرسیون مقادیر پیوسته محاسبه می‌شوند.

یادگیری نیمه‌نظارتی

زیرمجموعه‌ای از الگوریتم‌های یادگیری بانظارت هستند که نه تنها از داده‌های برچسب خورده استفاده می‌کنند، بلکه از داده‌های بدون برچسب نیز در محاسبه تابع نیز استفاده می‌کنند. یکی از مثال‌های این حوضه دسته‌بندی عکس افراد است. زمانی که صدها عکس از دوستان و آشنایان خود دارید و تنها چند عکس از هر نفر را برچسب می‌زنید و بقیه عکس‌ها را با استفاده از این الگوریتم برچسب می‌زنید. بسیاری از مسائل مطرح شده می‌توانند از این الگوریتم‌ها سود ببرند. برچسب‌زنی داده‌های ورودی الگوریتم‌ها می‌تواند بسیار هزینه‌بر باشد. در حالی که استفاده از این الگوریتم‌ها می‌تواند هزینه را بسیار کاهش دهد.

یادگیری بدون نظارت

این دسته از الگوریتم‌ها نیازی به داده‌های برچسب خورده ندارند. الگوریتم‌های یادگیری بدون نظارت تلاش می‌کنند که ساختارهای پنهان در اطلاعات بدون برچسب را استخراج کنند. از معروف‌ترین الگوریتم‌هایی که در این دسته‌بندی قرار می‌گیرند الگوریتم K-means است. این الگوریتم سعی در گروه‌بندی داده‌ها برای اساس فاصله منطقی آنها از یکدیگر دارد.

یادگیری تقویتی

این الگوریتم‌ها مانند یک حلقه بسته هستند. عامل هوشمند عملی را بر روی محیط انجام می‌دهد و سپس عمل او را ارزیابی می‌شود. میزان پاداش این عمل به همراه وضعیت محیط در اختیار عامل قرار می‌گیرد و او براساس اطلاعات دریافتی یاد می‌گیرد که چه عملی را باید انجام دهد. بنظر می‌رسد این مجموعه از الگوریتم‌ها به رؤیای هوش مصنوعی جامع نزدیک‌تر هستند.

سؤال دوم

پیچیدگی بیش از حد مدل و داشتن تعداد پارامترهای خیلی زیاد از دلایل عمده بیش‌برازش هست. همچنین در صورتی که مقدار تابع هزینه در زمان محاسبه مدل، بسیار کم باشد و میزان خطا در پیش‌بینی داده‌های فاز آزمون بسیار بالا باشد می‌توان به بیش‌برازش مدل محاسبه شده شک کرد. تغییرات قابل توجه در اندازه پارامترها با اضافه شدن داده‌های جدید و یا استفاده از داده‌های متفاوت برای مرحله ساخت مدل نیز از دیگر نشانه‌های بیش‌برازش خواهد بود.

سؤال سوم

اگر P وکتوری از داده‌های پیش‌بینی شده باشد و Y وکتوری از داده‌های اصلی، خطا MSE به شکل زیر محاسبه می‌شود:

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - Y_i)^2$$

و همچنین خطا $RMSE$ به شکل زیر خواهد بود:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - Y_i)^2}{n}}$$

همانطور که از فرمول‌ها مشخص است، خطا MSE در مقابل داده‌های پرت و ناهنجار واکنش بیشتری نشان می‌دهد. در این فرمول اگر یکی از داده‌های از مدل پیش‌بینی شده فاصله زیادی داشته باشد خطا به شدت افزایش پیدا می‌کند. به همین دلیل مدل پیش‌بینی شده برای کاهش خطا به سمت داده‌های پرت بیشتر کشیده می‌شود. این مشکل در خطا $RMSE$ به دلیل استفاده از جذر خطا MSE کاهش پیدا کرده است. بنابراین برای مجموعه داده‌ای با داده پرت و ناهنجار زیاد، خطا $RMSE$ بهتر خواهد بود.

سؤال چهارم

در حالتی که گرادیان نزولی در حال اجرا بر روی دره‌ای با عمق بسیار کم و دیوارهایی با شیب بسیار زیاد باشد، بعد از چند بروزرسانی سرعت نزول بسیار کم خواهد شد. تکانه عملاً مقدار چرخه‌های پیشین در گرادیان نزولی را به آخرین چرخه‌ای که در حال بروزرسانی آنست اضافه می‌کند. بطور شهودی می‌توان اینگونه در نظر گرفت که با استفاده از تکانه چرخه‌های گرادیان نزولی اطلاعاتی از چرخه‌های پیشین را دارند و استفاده می‌کنند. حال اگر مقدار این تکانه زیاد باشد، جهت مقادیر بروزرسانی شده تغییر چندانی نمی‌کند و اگر مقدار آن کم باشد عملاً تکانه تأثیر چندانی نخواهد داشت.

سؤال پنجم

سؤال ششم

به طور معمول با افزایش داده‌های آموزش بیش‌برازش کاهش پیدا می‌کند. با فرض اینکه در مدل کنونی، با داده‌های آموزش کم، بیش‌برازش اتفاق افتاده باشد، با افزایش داده‌ها تعداد هزینه مدل بیش‌برازش شده از هزینه مدل ساده بیش‌تر خواهد شد و در نتیجه مدل ساده انتخاب می‌گردد. تنها حالتی که امکان دارد استفاده از داده‌های آموزش بیش‌تر، در ایجاد مدل نقش منفی داشته باشد زمانی است که داده‌های آموزش کوچک‌تر بهتر باشند ولی داده‌های آموزش بزرگ‌تر اختلالات³ بیش‌تری داشته باشد که استفاده از آن باعث می‌شود مدل ایجاد شده دقیق نباشد.

سؤال هفتم

روش اول Regularization:

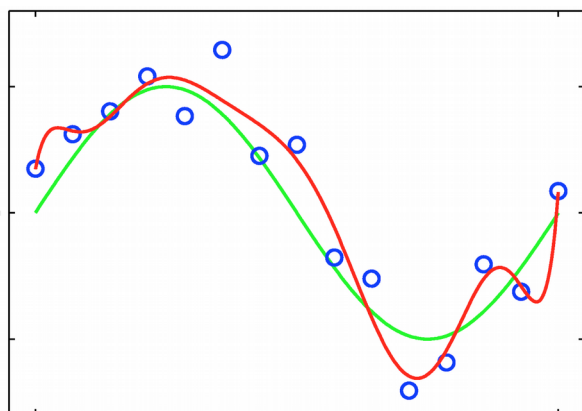
در این روش با در نظر گرفتن پنالتی برای ساختارها و پارامترهای پیچیده‌تر در مدل، آن را به سمت یک مدل ساده و هموار سوق می‌دهیم.

روش دوم Cross Validation:

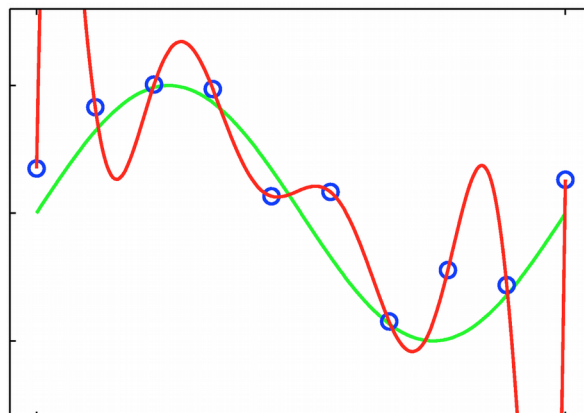
در این روش داده‌ها را به چند دسته تقسیم می‌کنیم. اگر داده‌ها به K دسته تقسیم شده باشند، $K-1$ دسته از آن‌ها برای داده‌های آموزش و یک دسته برای بررسی خطا مدل ساخته شده استفاده می‌شود. این روش K دفعه تکرار شده و هر کدام از دسته‌ها دقیقاً یک‌بار به عنوان دسته بررسی استفاده خواهند شد. در پایان داده‌های بدست آمده از این روش را با یکدیگر تلفیق می‌کنیم (به عنوان نمونه می‌توان از آن‌ها میانگین گرفت).

روش سوم استفاده از داده‌های خوب و کافی است:

اگر تعداد داده‌ها کم باشد و بخواهیم یک مدل بر روی آن‌ها پیدا کنیم، مدل‌های ما به راحتی دچار بیش‌برازش خواهند شد مانند شکل تصویر 1. در صورتی اگر داده‌های آموزش بیشتر باشند و اختلالات کمتری در آن‌ها باشد مدل ساخته شده کمتر مستعد بیش‌برازش خواهد بود مانند تصویر 2.



تصویر 2: داده‌های آموزش زیاد کمتر استعداد بیش‌برازش دارند.



تصویر 1: داده‌های آموزش کوچک مستعد بیش‌برازش

روش چهارم کاهش ویژگی⁴

زمانی که تعداد ویژگی‌ها زیاد باشد پیچیدگی مدل نیز بسیار بالا خواهد رفت. همچنین محاسبه نیز بیش‌تر خواهد شد. بهتر است ویژگی‌های مرتبط و آن‌هایی را که می‌توان از روی دیگر ویژگی‌ها محاسبه کرد حذف کنیم. همچنین استفاده از یک متخصص در آن زمینه نیز بسیار مفید خواهد بود.

روش‌های دیگر مانند تکانه، انتخاب خودکار ویژگی‌ها و ... وجود دارند.

سؤال هشتم

فایل AmirKhazaie-96131038-Question-8.ipynb در پیوست با استفاده از Jupyter پیاده‌سازی شده است.