



دانشکده مهندسی کامپیوتر

تمرین سری اول درس یادگیری ماشین

توضیحات

هدف از این تمرین آشنایی و تمرین مفاهیم تئوری و عملی می‌باشد. این تمرین شامل ۲ بخش می‌باشد. استفاده از هر منبعی برای پاسخگویی به سوالات ممکن می‌باشد و توصیه می‌شود موضوعات را بیشتر حد خواسته شده مطالعه کنید.

بخش اول شامل سوالات تئوری است و اگر بیان شد فرمول ریاضی را بررسی کنید باید فرمول را نوشته و ویژگی‌های تولید شده توسط هر قسمت را بیان کنید. در غیر این صورت نیازی به بررسی ریاضی نمی‌باشد، به صورت تجربی توضیح دهید.

بخش دوم سوالات برنامه نویسی می‌باشد. دقت کنید در بخش عملی بعد از گزارش نتایج هر قسمت نکات مورد توجه آن قسمت را بررسی و تحلیل کنید. قبل از گزارش نتایج نیز پارامترهای فرض شده و پیش فرض‌های انجام شده را بیان نمایید. کامنت گذاری کلی (فقط برای قسمت‌های مهم در حد چند کلمه) در کد الزامی می‌باشد. نمودارها باید تیتراشته باشند و محورها نام گذاری شده باشند.

ملاک اصلی ارزیابی شما گزارش شما می‌باشد و هرگونه کپی برداری از کار دیگران غیر مجاز است و جریمه خواهد شد. در اول هر پاسخ مشخص کنید به کدام سوال پاسخ می‌دهید. فایل‌های خود را در فایلی به نام StudentID_HW01 فشرده و بارگذاری کنید.

بخش اول

۱- مفاهیم زیر را توضیح دهید.

الف (یادگیری نظارتی (Supervised Learning)

ب (یادگیری نیم نظارتی (Semi-Supervised Learning)

پ (یادگیری بدون نظارت (UnSupervised Learning)

ت (یادگیری تقویتی (Reinforcement Learning)

ث (رگرسیون (Regression)

ج (یادگیری برخط (Online Learning)

چ (یادگیری فعال (Active Learning)

ح (یادگیری انتقالی (Transfer Learning)

خ (دسته بندی (Classification)

د (یادگیری قیاسی (Deductive Learning)

ذ (یادگیری استنتاجی (Inductive Learning)

ر (درون یابی (interpolation)

ز (برون یابی (extrapolation)

ژ (بیش برآزش (over-fitting)

۲- شباهت و تفاوت دسته بندی را با رگرسیون بررسی کنید. چگونه میتوان هرکدام را به دیگری تبدیل کرد.

۳- همبستگی بین ویژگی‌ها به چه معنی است و چگونه میتوان آن را تشخیص داد (فرمول ریاضی را بررسی کنید). مزایا و معایب همبستگی بین ویژگی‌ها را بررسی کنید.

۴- صفحه‌ی ۱۳ از پیوست ۱ را (روش Locally Weighted Regression) مطالعه کنید. تفاوت و شباهت این روش را با رگرسیون بررسی کنید.

۵- روش گرادیان نزولی را با روش معادله‌ی نرمال مقایسه کنید. شرایط کاربرد هرکدام را بیان کنید.

۶-

الف (تعریف بایاس و واریانس در مدل چگونه انجام میشود (ریاضی و تجربی).

ب) در مدل‌ها چگونه محدوده‌ی مناسب برای این دو تعریف تعیین میشود.

پ) این دو تعریف (بایاس و واریانس) در فرمول مجموع مربعات خطا (MSE) چگونه ظاهر می شوند.

بخش اول

۷- دلیل انجام Regularization چیست و با افزایش ضریب آن در فرمول خطا کدام پارامترهای مدل (کدام θ_i ها) بیشتر کاهش می‌یابند.

۸- به صورت معمول داده‌ها را به ۳ دسته تقسیم می‌کنند، آموزش، آزمون و ارزیابی. هر کدام به چه منظوری استفاده می‌شود و ممکن است چند بار استفاده شود.

۹- سه معیار خطای MSE و MAE و RMSE را با هم مقایسه کنید. هر کدام برای چه شرایطی مناسب می‌باشد.

پیوست

مجموعه‌ی داده به نام Dataset را بارگذاری کنید.

۱- داده ها را رسم کنید.

۲- تابعی بنویسید که با دریافت داده ها و عدد درجه‌ی مورد نظر، نرخ یادگیری، تعداد قدم و ضریب رگولاریزیشن خطی/ منحنی با روش گرادیان نزولی با رگولاریزیشن بر روی داده ها برازش کند و آن را همراه داده ها رسم کند و خطای MSE را نیز گزارش کند. مقدار حدودی مناسب چهار پارامتر قدم، درجه، نرخ یادگیری و λ را بیابید و گزارش کنید.

الف) نمودار وزن های مدل را در هر قدم ثبت کنید و در نموداری نشان دهید (برای با و بدون رگولاریزیشن). این دو نمودار چه تفاوتی دارند.

ب) دو پارامتر قدم و درجه را ثابت بگیرید و برای پارامتر نرخ یادگیری ۳ مقدار با فاصله‌ی مناسب امتحان کنید و نتیجه را با داده ها رسم کنید و نتیجه را بررسی کنید.

پ) دو پارامتر نرخ یادگیری و درجه را ثابت بگیرید و برای پارامتر قدم ۳ مقدار با فاصله‌ی مناسب امتحان کنید و نتیجه را با داده ها رسم کنید و نتیجه را بررسی کنید.

ت) دو پارامتر قدم و نرخ یادگیری را ثابت بگیرید و برای پارامتر درجه ۳ مقدار با فاصله‌ی مناسب امتحان کنید و نتیجه را با داده ها رسم کنید و نتیجه را بررسی کنید.

ث) به نظر شما به جای ویژگی استفاده شده از چه ویژگی با چه خاصیت ای استفاده میشد می‌توانست باعث جواب بهتری بشود؟ (در تعداد قدم مشخصی نمودار بهتری تولید کند نسبت به نمودار برازش شده‌ی شما)

۳- نمودار وزن های مدل را برای مقادیر مختلف λ نمایش دهید.

۴- تابعی بنویسید که با دریافت داده ها و عدد درجه‌ی مورد نظر خطی / منحنی به روش معادله‌ی نرمال بر روی داده ها برازش کند. سه عدد درجه‌ی مختلف با فاصله‌ی مناسب را امتحان کنید و نمودار خط را همراه با داده ها رسم کنید و نتایج را بررسی کنید.

- 1- <http://cs229.stanford.edu/notes2019fall/cs229-notes1.pdf>