



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر و فناوری اطلاعات
تمرین سری اول درس یادگیری ماشین

پاییز ۱۳۹۶

سؤال ۱

مفاهیم زیر را تعریف کرده و برای هر کدام توضیح مختصری بدهید.

- یادگیری با نظارت
- یادگیری نیمه نظارتی
- یادگیری بدون نظارت
- یادگیری تقویتی
- یادگیری برخط^۱

سؤال ۲

فرض کنید یک مدل بر روی یک مجموعه داده قرار است آموزش ببیند (مثلاً یک نمودار را بر روی چندین نقطه برازش کنیم). بدین منظور از ۸۰٪ داده‌ها برای فاز آموزش و از ۲۰٪ باقی‌مانده داده‌ها برای فاز آزمون استفاده می‌کنیم. چگونه متوجه بشویم که مدل دچار بیش برازش شده است یا خیر؟

سؤال ۳

خطای MSE و RMSE را تعریف کنید. در دیتاستی با داده پرت و ناهنجار زیاد، استفاده از کدام یک بهتر است؟ چرا؟

سؤال ۴

گاهی در هنگام استفاده از روش گرادیان نزولی، از اثر تکانه^۲ استفاده می‌شود. مزیت و دلیل استفاده از این اثر را شرح دهید. تکانه زیاد یا کم چه مشکلی پیش می‌آورد؟ (هر دو حالت بررسی شود)

سؤال ۵

مقدار θ_0 در بردار θ چه نقشی دارد؟ چرا θ_0 در قسمت regularization تابع هزینه، لحاظ نمی‌شود؟

سؤال ۶

با افزایش داده‌های آموزشی، احتمال بروز بیش برازش چه تغییری می‌کند؟ توضیح دهید.

^۱ online

^۲ momentum

سؤال ۷

چهار روش برای جلوگیری از بروز بیش برآزش ذکر کنید. برای هر مورد توضیح مختصری نیز بدهید.

سؤال ۸

داده‌های موجود در فایل اکسل data را در نظر بگیرید. ستون A مقادیر x و ستون B مقادیر y متناظر را نشان می‌دهد. برای تمامی قسمت‌های سؤال، ۸۰٪ داده‌ها را برای آموزش و ۲۰٪ مابقی را برای آزمون استفاده کنید.

الف) با استفاده از روش گرادیان نزولی و به ازای درجه‌های ۳ و ۵ و ۷ یک نمودار بر روی نقاط برآزش کنید. برای هر درجه یادشده، ۱۰۰، ۱۰۰۰، ۱۰۰۰۰ بار الگوریتم را تکرار کنید. نمودار را رسم کرده و مقدار خطا (MSE) (برای هر دو فاز آموزش و آزمون) و اندازه قدم (آلفا) را گزارش کنید.

خروجی موردنظر: تصویر نمودار برآزش شده بر روی نقاط به روش گرادیان نزولی، گزارش مقدار خطا (MSE) (برای هر دو فاز آموزش و آزمون به صورت نمودار؛ محور افقی دفعات تکرار، محور عمودی خطا) و اندازه قدم (α)، به ازای درجات ۳ و ۵ و ۷

ب) سؤال قبل را به روش معادله نرمال حل کنید. (ضریب λ را در نظر نگیرید)

خروجی موردنظر: تصویر نمودار برآزش شده بر روی نقاط به روش معادله نرمال، گزارش مقدار خطا (MSE) (برای هر دو فاز آموزش و آزمون به صورت نمودار؛ محور افقی درجه، محور عمودی خطا) به ازای درجات ۳ و ۵ و ۷

ج) به ازای درجه ۷، و به ازای مقدار λ برابر با ۵ و ۵۰ و ۵۰۰ و با استفاده از معادله نرمال، نمودار را بر روی نقاط برآزش کرده و رسم کنید. مقدار خطا (MSE) (برای هر دو فاز آموزش و آزمون) و بردار ضرایب θ را گزارش کنید. تغییر λ ، چه تأثیری بر اندازه بردار θ دارد؟

خروجی موردنظر: تصویر نمودار برآزش شده بر روی نقاط به روش معادله نرمال و درجه ۷ و به ازای مقدار λ برابر با ۵ و ۵۰ و ۵۰۰، گزارش مقدار خطا (MSE) (برای هر دو فاز آموزش و آزمون به صورت نمودار؛ محور افقی λ ، محور عمودی خطا) و بردار ضرایب θ و تأثیر λ بر روی اندازه بردار θ به ازای حالات ذکرشده.

توضیحات تمرین:

۱- شما باید سورس کد خود به همراه مستندات (پاسخ سؤال‌ها و نتایج پیاده‌سازی که خواسته شده است) را در قالب یک فایل *RAR* که نام فایل *xxxxxx_hw1* که *xxxxxx* شماره دانشجویی شما است، تحویل بدهید.

۲- پیاده‌سازی با متلب یا پایتون (نسخه 2.7) باید انجام شود.

۳- مجاز به استفاده از هیچ کتابخانه آماده‌ای نیستید.

۴- در صورت هرگونه سؤال یا ابهام به idin47@gmail.com ایمیل بزنید.

۵- مهلت انجام این تمرین تا ساعت ۲۳:۵۵ روز ۳۰م مهر است.