



دانشکده فنی دانشگاه تهران

دانشکده برق و کامپیوتر

تمرین ۱ سیستم‌های هوشمند

رایانامه

hesam.77s@gmail.com

yasamin.1998@gmail.com

طراحان:

امیرحسام سلیم نیا

یاسمین نیکنام

نیم سال اول ۱۳۹۹-۱۴۰۰

دانشجویان عزیز، قبل از پاسخ‌گویی به سوالات به نکات زیر توجه کنید:

۱. شما باید کدها و گزارش خود را با الگو IS_HW1_StudentNumber.zip در محل تعیین شده آپلود کنید.

۲. گزارش کار شما نیز از معیارهای ارزیابی خواهد بود، در نتیجه زمان کافی برای تکمیل آن اختصاص دهید.

۳. شما می‌توانید سوالات خود را از طریق ایمیل طراحان تمرین بپرسید.

۱. تابع زیر را در نظر بگیرید.

$$f(x) = 3x_1^2 + 2x_2^2 - 3x_1x_2 + 4x_1^3 + x_1^4$$

تمام نقاط ایستای تابع را بدست آورده و نوع آن را مشخص کنید.

۲. تابع زیر را در نظر بگیرید.

$$f(x) = 3x_1^2 + 2x_1 + 8x_2 + 4x_2^2$$

می‌خواهیم به کمک روش گرادیان نزولی مقدار کمینه‌ی محلی تابع $f(x)$ را محاسبه کنیم.

(آ) اگر نقطه‌ی شروع $(0, 0)$ باشد، جهت گرادیان نزولی در این نقطه را محاسبه نمایید.

(ب) با جست و جوی خطی مقدار بهینه طول گام α را با روش تحلیلی محاسبه نمایید.

(ج) بار دیگر مقدار بهینه‌ی طول گام را از روش *Armijo* محاسبه نمایید و با یکدیگر مقایسه کنید.

(د) الگوریتم‌های فوق را با نقطه‌ی شروع به روزسانی شده تکرار نمایید.

۳. از کاربردهای مهم روش‌های بهینه‌سازی، استفاده آن‌ها در طبقه‌بندها است. در این بخش می‌خواهیم به کمک ماشین بردار پشتیبان (SVM) به دسته‌بندی مجموعه داده‌ی MNIST بپردازیم. این مجموعه داده شامل ۱۰ کلاس است که هر کلاس نماینده یکی از اعداد ۰ تا ۹ است. هر عدد به صورت یک سطر ذخیره شده است که سایز آن 28×28 است. یکی از راه‌های طبقه‌بندی داده‌های مسائل Multiclass استفاده از روش one vs all است که در آن می‌بایست به تعداد کلاس‌های مسئله طبقه‌بند جداگانه‌ای آموزش دهیم. بدین صورت که هر طبقه‌بند تعیین میکند که هر داده به کلاس متناظر تعلق دارد یا خیر. در صورتی که داده‌ای عضو تنها یکی از این ۱۰ کلاس تشخیص داده شود، آن داده را متعلق به این کلاس می‌دانیم. تابع هزینه‌ای که برای این مسئله در نظر گرفته شده است به صورت زیر است و می‌خواهیم آن را به دو روش Stochastic Gradient Descent و Batch Gradient Descent کمینه کنیم. لازم به ذکر است که در تابع هزینه‌ی داده شده Hyperparameter ای وجود دارد که مانع overfitting شود که مقدار آن به کمک روش Cross-validation انتخاب می‌گردد.

$$\mathcal{L} = \lambda \|w\|^2 + \frac{1}{n} \sum_{i=0}^n \max(0, 1 - y_i(w x_i - b)) \quad (۱)$$

نکات مهم جهت پیاده‌سازی:

- (آ) مقادیر اولیه پارامتر w را یک عدد تصادفی از توزیع گاوسی با میانگین ۰ و واریانس ۰.۰۱ در نظر بگیرید.
- (ب) جهت تعیین مقدار پارامتر λ لازم است از روش Cross-validation استفاده کنید، به این صورت که به ازای هر یک از مقادیر پیشنهادی زیر برای λ الگوریتم مذکور را انجام داده و میانگین دقت‌های بدست آمده به ازای هر مقدار λ با یکدیگر مقایسه کرده و λ ای که بیشترین دقت را به خود اختصاص دهد را انتخاب کنید. جهت سادگی محاسبه و انتخاب λ تنها بر اساس یکی از کلاس‌ها کافی است. (بهینه‌سازی به کمک یکی از روش‌های SGD و یا BGD کافی است)

$$\lambda = \{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1\}$$

- (ج) پس از تعیین بهترین مقدار λ ، به کمک تمام داده‌های train دو مدل را به کمک روش‌های SGD و BGD آموزش داده و سپس دقت نهایی مدل بدست آمده را به کمک داده‌های test محاسبه نمایید. همچنین لازم است تا نمودار loss بر حسب تعداد تکرار (epoch) و همچنین دقت مدل‌ها روی داده‌های train و test بر حسب تعداد تکرار (epoch) در یک نمودار رسم کنید. (مقدار طول گام را ثابت در نظر بگیرید)

- (د) بار دیگر بخش (ج) را با طول گام کاهشی تکرار نمایید. (شیوه‌ی کاهش گام به انتخاب خودتان است)

۴. یکی دیگر از الگوریتم‌های کلاسیک طبقه‌بندی الگوریتم KNN است که به طور کلی جهت طبقه‌بندی (classification) و بدست آوردن توزیع (distribution) استفاده می‌گردد. در این تمرین به کاربرد اول خواهیم پرداخت.

دادگانی در اختیار شما قرار گرفته است، که شما می‌بایست با استفاده از داده‌های train و الگوریتم KNN برچسب داده‌های test را پیش‌بینی نمایید.

(آ) به ازای مقادیر $K = \{2, 5, 10, 50\}$ داده‌های test را طبقه‌بندی کنید. از معیار اقلیدسی (Norm 2) برای محاسبه فاصله استفاده کنید.

(ب) با توجه به مشاهدات خود از قسمت پیش، نظراتان را در خصوص تاثیر مقادیر بسیار بزرگ و مقادیر بسیار کوچک K بر طبقه‌بندی بیان کنید.

(ج) در الگوریتم KNN از معیارهای مختلفی برای محاسبه فاصله استفاده می‌شود که می‌توان به Norm 1 و Norm 2 اشاره کرد. در ادامه به ازای بهترین K بدست آمده در بخش قبل، دو معیار فاصله‌ای که در ادامه آمده است را امتحان کنید.

$$d_1(x, y) = \max |x_i - y_i|, \quad d_2 = \sum_{i=1}^n |x_i - y_i|$$

(د) در صورتی که دقت طبقه‌بندی پایین است، علت آن را بیان کنید و راهی برای اصلاح آن بیان کنید.

موفق باشید