



به نام او

تکلیف طراحی سیستمهای هوشمند

دانشگاه تحصیلات تکمیلی علوم پایه زنجان - بهار ۹۸

سوال اول- مسئله رگرسیون خطی را در نظر بگیرید که در آن نمونه های آموزشی به صورت متفاوت وزن دار شده اند. فرض کنید که هدف کمینه کردن تابع هدف زیر می باشد:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m w^{(i)} (\theta^T x^{(i)} - y^{(i)})^2$$

الف) نشان دهید که $J(\theta)$ می تواند به صورت زیر نوشته شود.

$$J(\theta) = (X\theta - \bar{y})^T W (X\theta - \bar{y})$$

که W بیانگر ماتریس قطری است و X و \bar{y} نیز برطبق نمادهایی است که در کلاس تعریف گردیده اند. توضیح دهید که ماتریس W بیانگر چیست؟

ب) پاسخ معادله بالا را با استفاده از معادله نرمال به دست آورید.

ج) گزینه های زیر را با استفاده از مجموعه داده های آموزشی که همراه با تکلیف فرستاده می گردد، حل نمایید. مجموعه داده های آموزشی شامل دو فایل `features.txt` و `label.txt` می باشد که به ترتیب هر سطر فایل شامل بردار ویژگی ها و برچسب یک نمونه آموزشی می باشد.

۱) رگرسیون خطی ساده را بر روی مجموعه داده پیاده سازی نمایید. سپس مدل حاصل (معادله خط به دست آمده) را همراه با نمونه های آموزشی در یک شکل ترسیم نمایید.

۲) رگرسیون خطی وزن دار شده را بر روی مجموعه داده پیاده سازی نمایید. سپس مدل حاصل (معادله خط به دست آمده) را همراه با نمونه های آموزشی در یک شکل ترسیم نمایید. به هنگام ارزیابی تابع از معادله زیر برای تنظیم وزن ها استفاده نمایید.

$$w^{(i)} = \exp\left(-\frac{(x - x^{(i)})^2}{2\tau^2}\right)$$

مقدار τ را به ۰.۸ تنظیم نمایید.

۳) در قسمت دوم مقدار پارامتر τ را به ترتیب به مقادیرهای ۰.۱، ۰.۳، ۰.۵ و ۱.۰ تنظیم نمایید. سپس توضیح دهید با تغییر مقدار پارامتر مذکور مدل های آموزش دیده چه تغییری می یابد

سوال دوم- در این سوال نسخه وزن دار شده logistic regression مورد بررسی قرار می گیرد که در آن نمونه های آموزشی براساس نمونه آزمایشی وزن دار می گردند. تابع هزینه در این حالت به صورت زیر تعریف می گردد:

$$J(\theta) = -\frac{\lambda}{2} \theta^T \theta + \sum_{i=1}^m w^{(i)} \left[t^{(i)} \log y(x^{(i)}) + (1 - t^{(i)}) \log(1 - y(x^{(i)})) \right]$$

که $-\frac{\lambda}{2}\theta^T\theta$ ترم نرمالیزاسیون می‌باشد که برای عملکرد بهتر روش لازم می‌باشد. در این مسئله مقدار پارامتر λ را 0.0001 تنظیم نمایید.

الف) بردار گرادیان تابع هزینه را به صورت فرم بسته بیابید.

ب) گزینه های زیر را با استفاده از مجموعه داده های آموزشی که همراه با تکلیف فرستاده می گردد، حل نمایید. مجموعه داده های آموزشی شامل دو فایل features_r.txt و label_r.txt می باشد که به ترتیب هر سطر فایل شامل بردار ویژگی ها و برچسب یک نمونه آموزشی می باشد.

۱) رگرسیون لاجستیک ساده را بر روی مجموعه داده پیاده سازی نمایید و دقت حاصل را گزارش نمایید. معیار دقت برابر با تعداد نمونه های تستی که به درستی پیش بینی شده اند بر روی تعداد کل نمونه های آموزشی است.

۲) رگرسیون لاجستیک وزن دار شده را بر روی مجموعه داده پیاده سازی نمایید. به هنگام ارزیابی تابع از معادله زیر برای تنظیم وزن ها استفاده نمایید.

$$w^{(i)} = \exp\left(-\frac{(x - x^{(i)})^2}{2\tau^2}\right)$$

مقدار τ را به 0.8 تنظیم نمایید. معیار دقت این روش را با روس رگرسیون لاجستیک ساده مقایسه نمایید.

۳) در قسمت دوم مقدار پارامتر τ را به ترتیب به مقادیرهای 0.001 ، 0.005 ، 0.01 ، 0.5 ، 1 و 5 تنظیم نمایید. سپس توضیح دهید با تغییر مقدار پارامتر مذکور فرضیه های حاصل چه تغییری می یابد.