



یادگیری ماشین برای بیوانفورماتیک

بهار ۱۴۰۲

استاد: علی شریفی زارچی

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

تاریخ برگزاری: ۹ خرداد

کوئیز ۳

سوالات (۱۰۰ نمره)

۱. (۶۰ نمره) پاسخ کوتاه

به سوالات زیر به صورت کوتاه پاسخ دهید:

- به چه دلیل استفاده از تابع sigmoid به عنوان تابع فعالسازی مستعد مشکل Vanishing Gradients است؟
- در صورتی که وزن اولیه تمام نورون های یک شبکه را با مقداری یکسان و ثابت مقداردهی کنیم، چه مشکلی در آموزش مدل بوجود خواهد آمد؟
- تفاوت بین gradient clipping در دو حالت قطع کردن گرادیان با مقدار (clipping by value) و قطع کردن با اندازه (clipping by norm) را توضیح داده و به طور خلاصه علت بهتر بودن استفاده از حالت دوم در اکثر موارد را ذکر کنید.
- دو روش Stochastic Gradient Descent و mini-batch Gradient Descent را از نظر بایاس و واریانس گرادیان با یکدیگر مقایسه نمایید.
- شما برای آموزش مدل خود از روش Batch Gradient Descent برای بهینه سازی تابع هزینه استفاده می کنید اما مشاهده می نمایید که تابع هزینه شما برای داده های آموزش خطای زیادی دارد. پس از بررسی کد خود متوجه می شوید که دیتاهای آموزش را shuffle نکرده اید. آیا shuffle کردن داده های آموزش این مشکل را رفع میکند؟ به طور مختصر دلیل خود را توضیح دهید.

۲. (۶۰ نمره) Weighted Cross-Entropy Loss

شما در حال آموزش یک مدل طبقه بندی دو کلاسه با استفاده از یک شبکه عصبی ساده هستید که در آن نمونه های کلاس یک به مراتب بیشتر از نمونه های کلاس صفر هستند. شبکه چندلایه زیر را در نظر بگیرید:

$$z_1 = W_1 x^{(i)} + b_1$$

$$a_1 = \text{ReLU}(z_1)$$

$$z_2 = W_2 a_1 + b_2$$

$$\hat{y}^{(i)} = \sigma(z_2)$$

$$L^{(i)} = \alpha * y^{(i)} * \log(\hat{y}^{(i)}) + \beta * (1 - y^{(i)}) * \log(1 - \hat{y}^{(i)})$$

$$J = -\frac{1}{m} \sum_{i=1}^m L^{(i)}$$

ابعاد این متغیرها به صورت زیر است:

$$\hat{y}^{(i)} \in \mathbb{R}, y^{(i)} \in \mathbb{R}, x^{(i)} \in \mathbb{R}^{D_x \times 1}, W_1 \in \mathbb{R}^{D_{a_1} \times D_x}, W_2 \in \mathbb{R}^{1 \times D_{a_1}}$$

همچنین m نیز تعداد نمونه های آموزش است.

با توجه به این موضوع به سوالات زیر پاسخ دهید:

الف) پارامترهای α و β چه کاربردی میتوانند داشته باشند؟ به طور نسبی با توجه به توضیحات مسئله یک مقدار برای آنها پیشنهاد دهید.

ب) حاصل گرادیان $\frac{\partial J}{\partial W_1}$ را محاسبه نمایید. می توانید از نامگذاری برای مشتقات لایه های میانی استفاده کنید. (به طور مثال $\delta_1 = \frac{\partial J}{\partial \hat{y}}$ و ...)