Daria Bolgova

NE 795 Scientific Machine Learning Fall 2023

Homework 2

Fundamental Techniques of Machine Learning

Problem 1. **Logistic Regression.** *Using the notations given in the lecture slides, derive the partial derivative of* $LL(\beta)$ *w.r.t.* β_j*:*

$$\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{i=1}^{m} \left[y^{(i)} - \sigma\left(\beta^T x^{(i)}\right) \right] \cdot x_j^{(i)}.$$

The log-likelihood function is

$$LL(\beta) = \sum_{i=1}^{m} \left[y^{(i)} \cdot \ln\left(\sigma\left(\beta^T x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \cdot \ln\left(1 - \sigma\left(\beta^T x^{(i)}\right)\right) \right],$$

$$\beta = [\beta_0, \beta_1, \dots, \beta_d].$$

Sigmoid function is

$$\sigma(z) = \frac{1}{1 + e^{-z}}.$$

The log-likelihood function derives from:

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^{m} y^{(i)} \cdot \ln\left(\sigma\left(\beta^T x^{(i)}\right)\right) = \sum_{i=1}^{m} y^{(i)} \cdot \frac{\partial \ln\left(\sigma\left(\beta^T x^{(i)}\right)\right)}{\partial \sigma(\beta^T x^{(i)})} \cdot \frac{\partial \sigma\left(\beta^T x^{(i)}\right)}{\partial (\beta^T x^{(i)})} \cdot \frac{\partial \beta^T x^{(i)}}{\partial \beta_j} =$$

$$= \sum_{i=1}^{m} y^{(i)} \cdot \frac{1}{\sigma(\beta^T x^{(i)})} \cdot \frac{\partial \left(\frac{1}{1 + e^{-\beta^T x^{(i)}}} \right)}{\partial (\beta^T x^{(i)})} \cdot \frac{\partial \beta^T x^{(i)}}{\partial \beta_j} =$$

$$= \sum_{i=1}^{m} y^{(i)} \cdot \left(e^{-\beta^T x^{(i)}} + 1\right) \cdot \frac{e^{-\beta^T x^{(i)}}}{\left(e^{-\beta^T x^{(i)}} + 1\right)^2} \cdot x^{(i)} = \sum_{i=1}^{m} y^{(i)} \cdot \frac{e^{-\beta^T x^{(i)}}}{\left(e^{-\beta^T x^{(i)}} + 1\right)} \cdot x^{(i)};$$

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^{m} \left[\left(1 - y^{(i)}\right) \cdot \ln\left(1 - \sigma\left(\beta^T x^{(i)}\right)\right) \right] =$$

$$= \frac{\partial}{\partial \beta_j} \sum_{i=1}^{m} \left[\ln\left(1 - \sigma\left(\beta^T x^{(i)}\right)\right) - y^{(i)} \cdot \ln\left(1 - \sigma\left(\beta^T x^{(i)}\right)\right) \right];$$

$$\frac{\partial}{\partial\beta_j}\sum_{i=1}^{m}\ln\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)=\sum_{i=1}^{m}\frac{\partial\ln\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)}{\partial\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)}\cdot\frac{\partial\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)}{\partial\left(\beta^T x^{(i)}\right)}\cdot\frac{\partial\beta^T x^{(i)}}{\partial\beta_j}=$$

$$=\sum_{i=1}^{m}\frac{1}{1-\sigma(\beta^T x^{(i)})}\cdot\frac{-e^{-\beta^T x^{(i)}}}{\left(e^{-\beta^T x^{(i)}}+1\right)^2}\cdot x^{(i)}=\sum_{i=1}^{m}\frac{1+e^{-\beta^T x^{(i)}}}{e^{-\beta^T x^{(i)}}}\cdot\frac{-e^{-\beta^T x^{(i)}}}{\left(e^{-\beta^T x^{(i)}}+1\right)^2}\cdot x^{(i)}=$$

$$=\sum_{i=1}^{m}-\frac{1}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)};$$

$$\frac{\partial}{\partial\beta_j}\sum_{i=1}^{m}-y^{(i)}\cdot\ln\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)=\sum_{i=1}^{m}\frac{y^{(i)}}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)};$$

$$\frac{\partial LL(\beta)}{\partial\beta_j}=\frac{\partial}{\partial\beta_j}\sum_{i=1}^{m}\left[y^{(i)}\cdot\ln\left(\sigma\left(\beta^T x^{(i)}\right)\right)+\left(1-y^{(i)}\right)\cdot\ln\left(1-\sigma\left(\beta^T x^{(i)}\right)\right)\right]=$$

$$=\sum_{i=1}^{m}y^{(i)}\cdot\frac{e^{-\beta^T x^{(i)}}}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}+\sum_{i=1}^{m}-\frac{1}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}+\sum_{i=1}^{m}\frac{y^{(i)}}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}=$$

$$=\sum_{i=1}^{m}\frac{y^{(i)}\cdot e^{-\beta^T x^{(i)}}-1+y^{(i)}}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}=\sum_{i=1}^{m}\frac{y^{(i)}\cdot\left(e^{-\beta^T x^{(i)}}+1\right)-1}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}=$$

$$=\sum_{i=1}^{m}y^{(i)}-\frac{1}{e^{-\beta^T x^{(i)}}+1}\cdot x^{(i)}=\sum_{i=1}^{m}\left[y^{(i)}-\sigma\left(\beta^T x^{(i)}\right)\right]\cdot x_j^{(i)}.$$