Lecture X.Y Topic

wbg231

December 2022

1 introduction

- the posterior can be written as $P(w|\mathcal{D}) \propto P(\mathcal{D}|w)P(w) = \mathcal{L}_{\mathcal{D}}(w)P(w)$
- so the likelihood of our data set with parameter w is $\mathcal{L}_{\mathcal{D}}(w) = P(y_1...y_n|x_1...x_n, w)$
- under the assumptions of logistic regression this can further be broken down as $\mathcal{L}_{\mathcal{D}}(w) = P(y_1...y_n|x_1...x_n, w) = P(y_1|x_1, w)...P(y_n|x_n, w) = \Pi_{i=1}^{n} P(y_i|x_i, w)^n = P(y=1)^{n_p} P(y=-1)^{n_n}$ where n_n, n_p are the number of observations with positive and negative labels respectively
- we know that the log is a monotonically increasing function and thus $\mathcal{L}_{\mathcal{D}}(w) \propto log(\mathcal{L}_{\mathcal{D}}(w)) = \ell(w)$
- further under out logistic assumptions we can write our logistic likelihood as $\ell(w) = \frac{1}{2}\Sigma_{i=1}^{n}(1+y^i)log(P(y_1 = 1|x, w)) + (1-y^i)log(P(y_1 = -1|x, w)) = \frac{1}{2}\Sigma_{i=1}^{n}(1+y^i)log(\frac{1}{1-e^{-w^t x^i}}) + (1-y^i)log(1 - \frac{1}{1+e^{-w^t x^i}})$
- then assuming -1 is a valid constant of proportionality we can write $P(w|\mathcal{D}) \propto P(\mathcal{D}|w)P(w) \propto -(\frac{1}{2}\Sigma_{i=1}^{n}(1+y^i)log(\frac{1}{1-e^{-w^t x^i}}) + (1-y^i)log(1 - \frac{1}{1+e^{-w^t x^i}}))P(w)$
- and thus have written our posterior in terms of the negative log likelihood of our data set given the parameter and the prior distribution of our parameter w