Machine Learning: Project 1 (Addendum)

Olivier Stähli
olivier.staehli@epfl.ch

Ivan Bioli
ivan.bioli@epfl.ch

Fabio Matti
fabio.matti@epfl.ch

1 Logistic regression for labels in $\{-1, 1\}$

We start from the standard negative log-likelihood from [1] for labels $y_n \in \{0, 1\}$:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{w}) &= \sum_{n=1}^{N} \ln\left[1 + \exp(\boldsymbol{x}_n^\mathsf{T}\boldsymbol{w})\right] - y_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w} \\
&= \sum_{n=1}^{N} \ln\left[\frac{1 + \exp(\boldsymbol{x}_n^\mathsf{T}\boldsymbol{w})}{\exp(y_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w})}\right] \\
&= \sum_{n=1}^{N} \ln\Big[\exp(-\underbrace{y_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w}}_{=\,0 \text{ if } y_n = 0}) \\
&\qquad + \exp(\underbrace{(1 - y_n)\boldsymbol{x}_n^\mathsf{T}\boldsymbol{w}}_{=\,\boldsymbol{x}_n^\mathsf{T}\boldsymbol{w} \text{ if } y_n = 0})\Big] \\
&= \sum_{n=1}^{N} \ln\left[1 + \exp(-\hat{y}_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w})\right] \qquad (1)
\end{aligned}$$

where in the last step we replaced y_n with $\hat{y}_n \in \{-1, 1\}$. The gradient of this expression is

$$\begin{aligned}
\nabla\mathcal{L}(\boldsymbol{w}) &= \sum_{n=1}^{N} \frac{\exp\left(-\hat{y}_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w}\right)}{1 + \exp\left(-\hat{y}_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w}\right)}(-y_n \boldsymbol{x}_n) \\
&= \sum_{n=1}^{N} \frac{-y_n \boldsymbol{x}_n}{1 + \exp\left(\hat{y}_n \boldsymbol{x}_n^\mathsf{T}\boldsymbol{w}\right)} \\
&= -\boldsymbol{X}^\mathsf{T}\left[\boldsymbol{y} \odot \frac{1}{1 + \exp\left(\hat{\boldsymbol{y}} \odot (\boldsymbol{X}\boldsymbol{w})\right)}\right] \qquad (2)
\end{aligned}$$

Here, $\odot$ denotes element-wise multiplication (the Hadamard product).

References

[1] *Jaggi M., Urbanke R., Khan M. E.* Machine Learning (CS-433): Lecture notes. 2021. https://github.com/epfml/ML_course.