MACHINE LEARNING 2 SUMMER 2020

- 1. Exercise, Solution to problem 1 -

Solve exercise 1 in the lecture notes.

Solution:

first let us repeat the task from the lecture notes:

$$\begin{aligned}
& \underset{\theta\in\mathbb{R}^m}{argmax} \quad p_{\Theta|X,Y}(\theta, x, y) \\
\overset{3.3}{=} & \underset{\theta\in\mathbb{R}^m}{argmax} \quad \prod_{s\in S} p_{Y_s|X_s,\Theta}(y_s, x_s, \theta) \prod_{v\in V} p_{\Theta_v}(\theta_v) \\
= & \underset{\theta\in\mathbb{R}^m}{argmax} \quad \sum_{s\in S} \log p_{Y_s|X_s,\Theta}(y_s, x_s, \theta) + \sum_{v\in V} \log p_{\Theta_v}(\theta_v) \qquad (3.9)
\end{aligned}$$

Substituting in (3.9) the linearization

$$\begin{aligned}
& \log p_{Y_s|X_s,\Theta}(y_s, x_s, \theta) \\
= \quad & y_s \log p_{Y_s|X_s,\Theta}(1, x_s, \theta) + (1 - y_s) \log p_{Y_s|X_s,\Theta}(0, x_s, \theta) \\
= \quad & y_s \log \frac{p_{Y_s|X_s,\Theta}(1, x_s, \theta)}{p_{Y_s|X_s,\Theta}(0, x_s, \theta)} + \log p_{Y_s|X_s,\Theta}(0, x_s, \theta) \qquad (3.10)
\end{aligned}$$

as well as (3.4) and (3.5) yields the form (3.11) below that is called the instance of the l_2-regularized logistic regression problem with respect to x, y and σ.

$$\underset{\theta\in\mathbb{R}^m}{argmin} \sum_{s\in S} \left(-y_s\langle\theta, x_s\rangle + \log\left(1 + 2^{\langle\theta, x_s\rangle}\right)\right) + \frac{\log e}{2\sigma^2}\|\theta\|_2^2 \qquad (3.11)$$

a) Derive (3.11) from (3.10) using (3.9), (3.4) and (3.5)
b) Is the objective function of (3.11) convex?

Solution a):

$$
\begin{aligned}
p_{Y_s|X_s,\Theta}(0, x_s, \theta) &= 1 - p_{Y_s|X_s,\Theta}(1, x_s, \theta) \\
&= 1 - \frac{1}{1 + 2^{-\langle\theta, x_s\rangle}} \\
&= \frac{2^{-\langle\theta, x_s\rangle}}{1 + 2^{-\langle\theta, x_s\rangle}} \quad (S.1) \\
&= \frac{1}{1 + 2^{+\langle\theta, x_s\rangle}} \quad (S.2)
\end{aligned}
$$

We now put (S.1) in the first term of (3.10) and (S.2) in the second term of (3.10):

$$
\begin{aligned}
& y_s \log \frac{p_{Y_s|X_s,\Theta}(1, x_s, \theta)}{p_{Y_s|X_s,\Theta}(0, x_s, \theta)} + \log p_{Y_s|X_s,\Theta}(0, x_s, \theta) \\
= \; & y_s \log \frac{1}{2^{-\langle\theta, x_s\rangle}} - \log(1 + 2^{+\langle\theta, x_s\rangle}) \\
= \; & + y_s \langle\theta, x_s\rangle - \log(1 + 2^{+\langle\theta, x_s\rangle})
\end{aligned}
$$

Using this in (3.9) and switching from argmax to argmin, which negates the term, results in (3.11) without the last term. In the last term $\frac{1}{\sigma\sqrt{2\pi}}$ is a constant factor which becomes an additive constant after the log and therefore does not change the *argmin*.

Solution b) (convexity of (3.11)):

- (3.11) is defined on $\mathbb{R}^m$, i.e. a convex set.
- A sum is convex if all summands are convex.
- First summand is linear, i.e. convex.
- Last summand is a sum of squares, hence convex.
- "A twice continuously differentiable function of several variables is convex on a convex set if and only if its Hessian matrix of second partial derivatives is positive semidefinite on the interior of the convex set."(copied from Wikipedia)
- We need to consider the Hessian matrix of the term $\log\left(1 + 2^{\langle\theta, x_s\rangle}\right)$:

$$h_{i,j} = \frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\log\left(1+2^{\langle\theta,x_s\rangle}\right) \tag{1}$$

$$= \frac{\partial}{\partial\theta_i}\frac{ln(2)\cdot x_{s,j}\cdot 2^{+\langle\theta,x_s\rangle}}{ln(2)\cdot(1+2^{+\langle\theta,x_s\rangle})} \tag{2}$$

$$= \frac{\partial}{\partial\theta_i}\frac{x_{s,j}}{(1+2^{-\langle\theta,x_s\rangle})}$$

$$= -\frac{x_{s,j}}{(1+2^{-\langle\theta,x_s\rangle})^2}\cdot(ln(2)\cdot(-x_{s,i})) \tag{3}$$

$$= \frac{ln(2)}{(1+2^{-\langle\theta,x_s\rangle})^2}\cdot(x_{s,i}x_{s,j}) \tag{4}$$

The first factor is a positive constant c. The second factor is the matrix

$$M = x_s \cdot x_s^T \tag{5}$$

M is positive semi-definite if for any y

$$y^T \cdot M \cdot y \geq 0 \tag{6}$$

Because the terms in a scalar product and factors in a product can be interchanged we get

$$y^T \cdot x_s \cdot x_s^T \cdot y \geq 0 \tag{7}$$

$$(y^T \cdot x_s)\cdot(x_s^T \cdot y) \geq 0 \tag{8}$$

$$\langle y, x_s\rangle^2 \geq 0 \quad \forall\ y \tag{9}$$

□