ML using Matlab, Sheet 1

Florian Wolf

Question 1

We visualize the (dis)advantages of the two options in a tabular

	advantages	disadvantages
scanned image in email	• unchanged / not compressed → "real" information	• possibly really large file size
OCR as txt in email	• way smaller file size - instant information usage for a machine • access from every system (one just needs a command line)	- loss and possible corruption of data due to scanning tool - possible no / bad result because of bad scan quality

Usage: • If one wants to process further the data via a machine respectively one only wants the contained data (and not all information) or only a low data transmission is available, one should use OCR

• If one has only a low scanner quality available or a possible loss of information is no option, then sending the scanned image is the way to go.

Question 2:

We list different options and shortly evaluate them

• Check labels manually by hand: This option provides the safest results but the effort is so high, that this is often impossible or a lot of people are needed.

- Increase scanning quality: Could be cost intensive.

- Define a measure to compare similar data points (eg. norm of the differences of two vectors containing image pixels). Here we have two options:
  - → check labels manually if the data is similar (using a threshold) and the labels differ
  - → give probability, using a classifier, that a data point is misslabeled (comparing multiple similar data points) with respect to the measure and decrease its weight in the calculation process (eg. in weighted logistic regression)

Question 3 Due to lack of notation, we assume $w, b \in \mathbb{R}$.

Proof:

linear regression:

$$\frac{\partial}{\partial w} L(w,b) = \frac{\partial}{\partial w}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(w x^{(i)} + b - y^{(i)}\right)^2\right) = \frac{1}{m}\sum_{i=1}^{m}\left(w x^{(i)} + b - y^{(i)}\right) x^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$\frac{\partial}{\partial b} L(w,b) = \frac{\partial}{\partial b}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(w x^{(i)} + b - y^{(i)}\right)^2\right) = \frac{1}{m}\sum_{i=1}^{n}\left(w x^{(i)} + b - y^{(i)}\right)\cdot 1$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)$$

logistic regression: Write ∂_w instead of $\frac{\partial}{\partial w}$

$$\partial_w\left(y^{(i)} \log h(x^{(i)})\right) = y^{(i)} \cdot \frac{1}{h(x^{(i)})} \cdot (-1) \cdot h(x^{(i)})^2 \cdot (-x^{(i)}) \exp(-w x^{(i)} + b)$$

$$= y^{(i)} \; h(x^{(i)}) \cdot (x^{(i)}) \exp(-(w x^{(i)} + b))$$

$$\partial_w\left(\log\left(1 - h(x^{(i)})\right)\right) = \frac{1}{1 - h(x^{(i)})} \cdot (-1) \cdot \left((-1) \cdot h(x^{(i)})^2 \cdot (-x^{(i)}) \exp(-(w x^{(i)} + b))\right)$$

$$= \frac{(-1)}{1 - h(x^{(i)})} h(x^{(i)})^2 \, x^{(i)} \exp(-(w x^{(i)} + b))$$

Therefore we obtain using the derivation rules for sums and the calculations above

$$\frac{\partial}{\partial w} L(w,b) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} h(x^{(i)})\, x^{(i)}\, e^{-(w x^{(i)}+b)} - \frac{1-y^{(i)}}{1-h(x^{(i)})} h(x^{(i)})^2 x^{(i)} e^{-(w x^{(i)}+b)}$$

$$= -\frac{1}{m} \sum_{i=1}^{m} h(x^{(i)})\, x^{(i)} e^{-(w x^{(i)}+b)} \cdot \left(y^{(i)} - (1-y^{(i)}) \frac{1}{\frac{1}{h(x^{(i)})}-1} \right)$$

$$= -\frac{1}{m} \sum_{i=1}^{m} h(x^{(i)})\, x^{(i)} e^{-(w x^{(i)}+b)} \cdot \underbrace{\left(y^{(i)} - (1-y^{(i)}) \cdot e^{w x^{(i)}+b} \right)}_{=\left(y^{(i)}\left(1+e^{w x^{(i)}+b}\right) - e^{w x^{(i)}+b}\right)}$$

$$\overset{(*)}{=} -\frac{1}{m} \sum_{i=1}^{m} h(x^{(i)})\, x^{(i)} \left(y^{(i)} \cdot \underbrace{\left(1 + e^{-(w x^{(i)}+b)} \right)}_{=1/h(x^{(i)})} - 1 \right)$$

minus into brackets.

$$= \frac{1}{m} \sum_{i=1}^{m} x^{(i)} \left(h(x^{(i)}) - y^{(i)} \right)$$

In an analogous way, we obtain

$$\partial_b \log(h(x^{(i)})) = \frac{1}{h(x^{(i)})} \cdot \partial_b h(x^{(i)}) = \frac{1}{h(x^{(i)})} \cdot (-1) \cdot h(x^{(i)})^2 e^{-(w x^{(i)}+b)} \cdot (-1)$$

$$= h(x^{(i)})^2 e^{-(w x^{(i)}+b)}$$

$$\partial_b \log(1-h(x^{(i)})) = \frac{1}{1-h(x^{(i)})} \cdot (-1) \cdot \underbrace{\partial_b (1-h(x^{(i)}))}_{=(-1)\cdot \partial_b h(x^{(i)})}$$

$$= \frac{1}{1-h(x^{(i)})} \partial_b h(x^{(i)})$$

$$\underset{\text{above}}{\overset{\text{calc. from}}{=}} \frac{1}{1-h(x^{(i)})} h(x^{(i)})^2 e^{-(w x^{(i)}+b)}$$

All in all, repeating the reshaping in (*) without the factor $x^{(i)}$, we obtain the following

$$\partial_b L(w,b) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} h(x^{(i)}) e^{-(w x^{(i)}+b)} + (1-y^{(i)}) \left(\frac{1}{1-h(x^{(i)})} h(x^{(i)})^2 e^{-(w x^{(i)}+b)} \right)$$

$$\overset{(*)}{=} \frac{1}{m} \sum_{i=1}^{m} h(x^{(i)}) - y^{(i)}$$

This was our initial claim. $\square$