به جای max likelihood از $\max \log(\ell\phi)$ استفاده می‌کنیم که همان $\min(-\log(\phi))$ است.

سوال ۱:

$$\phi_{sig}(w_1x_1 + w_2x_2) \rightarrow \frac{1}{1+e^{y_iw_1x_1}e^{y_iw_2x_2}} \rightarrow \min_w \log(1+e^{y_i(w_1x_1+w_2x_2)})$$

$$\rightarrow \frac{\partial(\phi_{sig})}{\partial w} = \left[\frac{y_ix_1}{1+e^{-y_iw^Tx_i}}, \frac{y_ix_2}{1+e^{-y_iw^Tx_i}}\right]$$

همان طور که دیدیم ∇f به مقادیر x وابسته است و چون در دیتای سوم بردار x صفر است ضرایب ما تغییر نمی‌کنند زیرا

$\frac{\partial sig}{\partial w}$ ما بردار $\begin{bmatrix}0\\0\end{bmatrix}$ است که اگر x برابر آن را از وزن‌ها کم کنیم وزن‌ها تغییری نمی‌کنند.

$$\phi_{sig}(w_0 + w_1x_1 + w_2x_2) = \frac{1}{1+e^{y_iw_0}e^{y_iw_1x_1}e^{y_iw_2x_2}} \rightarrow \min_w \log(1+e^{y_iw_0+y_iw_1x_1+y_iw_2x_2})$$

$$\rightarrow \frac{\partial(\phi_{sig})}{\partial w} = \left[\frac{1}{1+e^{-y_iw^Tx_i}}, \frac{y_ix_1}{1+e^{-y_iw^Tx_i}}, \frac{y_ix_2}{1+e^{-y_iw^Tx_i}}\right]$$

به علت وجود w_0 ی اول دیتای سوم ما هم در تغییر w نقش دارد چون فقط دو تا سطر دوم صفر می‌شوند و سطر اول عددی مخالف صفر است.

سوال ۲: الف)

$$w = [0, 0]$$

$$\hat{y} = w^Tx_1 = 0 \rightarrow w = \begin{bmatrix}0\\0\end{bmatrix} + 1\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

$$\hat{y} = w^Tx_2 = 2-1 = 1 \rightarrow w = \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}-1\\2\end{bmatrix}$$

$$\hat{y} = w^Tx_3 = 3-2 = 1 \rightarrow w = \begin{bmatrix}-1\\2\end{bmatrix} - \begin{bmatrix}-3\\-1\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$$

$$\hat{y} = -6+3 = -3 \rightarrow w = \begin{bmatrix}2\\3\end{bmatrix} + \begin{bmatrix}-3\\1\end{bmatrix} = \begin{bmatrix}-1\\4\end{bmatrix}$$

$$\hat{y} = w^Tx_1 = 3 \rightarrow \text{next}$$

$$\hat{y} = w^Tx_2 = -2-4 = -6 \rightarrow \text{next}$$

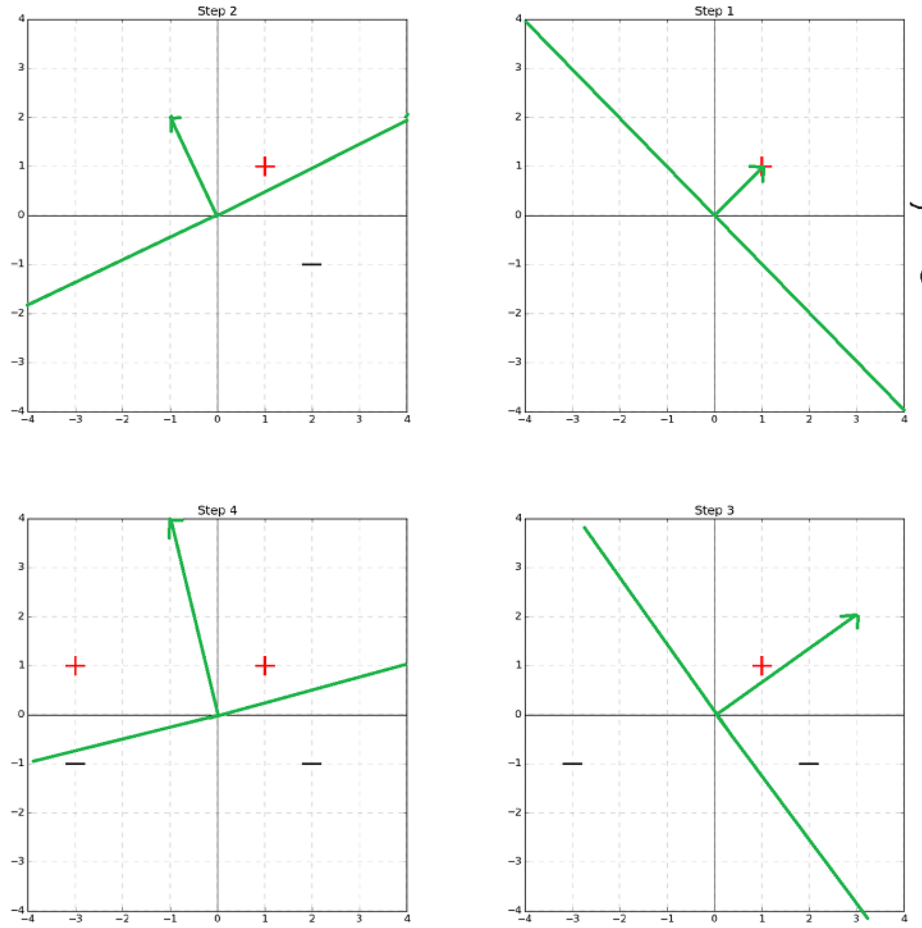
$$\hat{y} = w^Tx_3 = 3-4 = -1 \rightarrow \text{next}$$

$$\hat{y} = w^Tx_4 = 3+4 = 7 \rightarrow \text{next}$$

→ در ایتریشن دوم همگرا شد.

Scanned with CamScanner

s.a.m



برداری که کشیده شده w است و جهت آن همان ناحیه‌ی مثبت است و خط عمود بر آن جداکننده‌ی دو ناحیه‌ی مثبت منفی است.

ب) خیر الگوریتم پرسپترون تلاشی برای ماکسیمم کردن مرز بین داده ترین و تصمیم‌گیری نمی‌کند بلکه تلاش می‌کند داده‌ها را از هم جدا کند (تا زمانی ادامه می‌دهیم که $X . w . y$ مثبت شود که این یعنی داده‌ها جدا شوند و زمانی که به این نتیجه رسیدیم توقف می‌کنیم که این ممکن است بیشترین مارجین نباشد)

CS Scanned with CamScanner