Problem Set 2 {Supervised Learning II}

1. Logistic Regression: Training Stability

Goal: Develop your skills debugging machine Learning Algorithms.

→ Implementation of Logistic regression is Provided in src/pol_lr.py.

⇒ Two labeled dataset:

① data/ds1_a.txt

② data/ds1_b.txt

(a) * On dataset A, the training Converges in 2,78,103 iteration.

* On dataset B, the training didn't Converge until 50,000,00 iteration

↳ Maybe it never Converges

(b)
X, red → $y == 0$

o, blue → $y == 1$

learning_rate = 1

$\|\Delta\theta\| < 10^{-15}$ for Convergence

Observation

1) For dataset A ~~at algorithm~~ in the learning phase, θ tends to Converge to a fixed point

2) For dataset B, in the learning phase, θ tends to continuously increase.

⇒ If the dataset can be perfectly separated by a linear decision boundary, then the objective $\ell(\theta)$ can be arbitrarily maximized by just scaling the θ.

⤷ Multiplying θ by a factor does not change the decision boundary.

⤷ But Probability being assigned to each data point can be arbitrarily changed

$$h_\theta(A) = \frac{1}{1+e^{-\theta^T A}}$$

If $\theta \to k\theta$

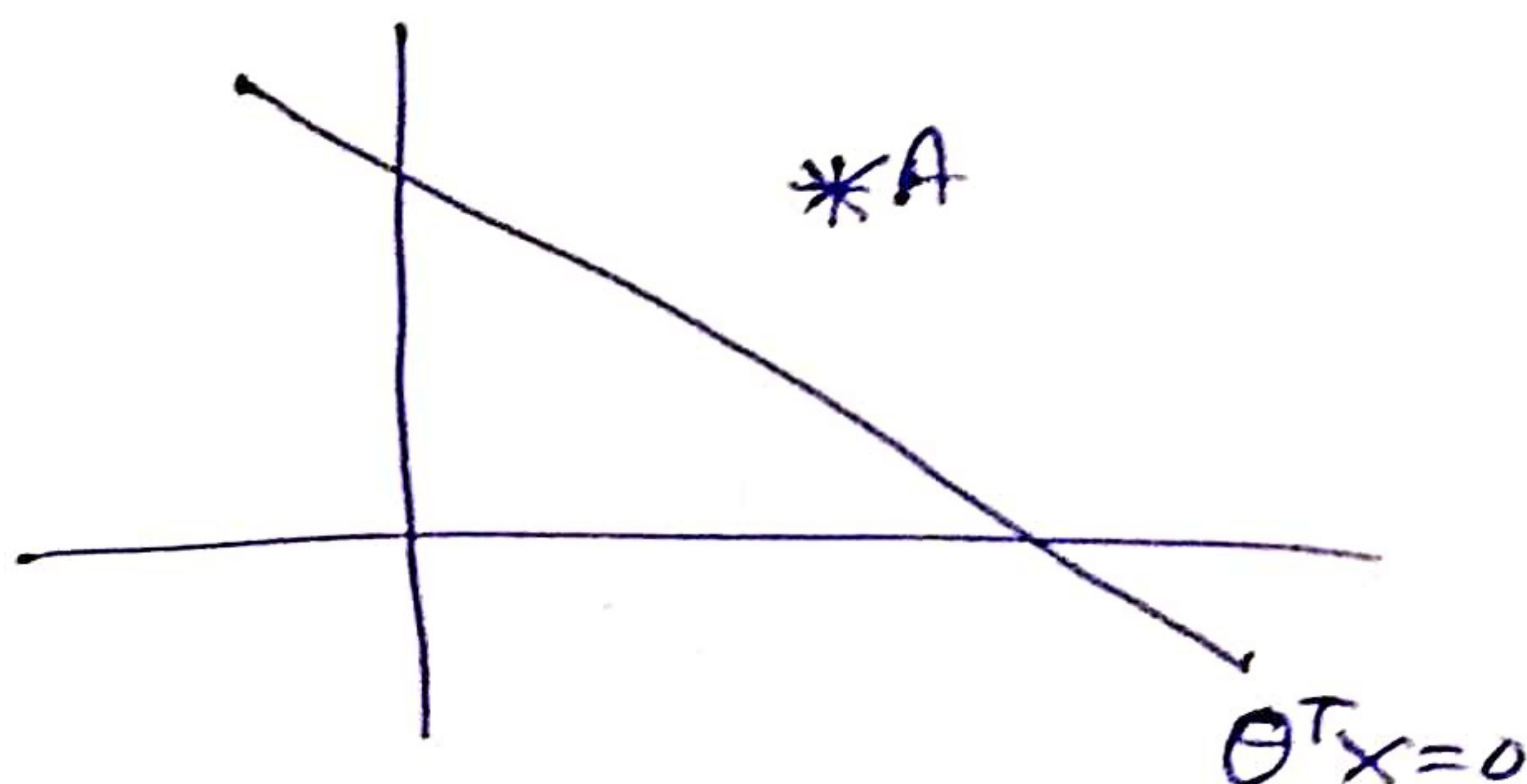$$h_\theta^*(A) = \frac{1}{1+e^{-k\theta^T A}}$$

$$h_\theta(A) < h_\theta^*(A)$$

⇒ as $k \to \infty$ $h_\theta^*(A) \to 1$

*A

$\theta^T x = 0$

◆ ds1_a.txt

⤷ Cannot be separated perfectly, so we are able to converge at some θ.

ds2_b.txt

⤷ Can be separated perfectly, so we are not able to converge at any θ.

©

(i) No, there is no maxima in $l(\theta)$

(ii) ~~No, there is no maxima in $l(\theta)$~~

(iii)(ii) Yes, when the learning rate is sufficiently small, the update to θ will be small, and it will be judge to be converged by the algorithm.

(iii) No, The dataset will still be linearly separable.

(iv) Yes, In this case there is a cost of In Choosing θ.

(v) Yes, this will very likely make the dataset (not) linearly separable

(d) No, SVM's objective is directly associated to geometric margin, which is independent with the scaling of θ.