

Homework 03

Zach Stecher

Due: 10/11/16

Problem 1) Run multiple random experiments using supplied data generation method and explain your observations about what works best in terms of E_{out} and the amount of computation required.

(a) The Pocket algorithm, starting from $w = 0$.

The pocket algorithm took a rather long time to run to completion, even compared to the original Perceptron. However, in instances where a clean solution was impossible, it managed to pick out the best of its guesses. In terms of computing resources needed and time to run, it seems to lag behind, but makes up for it with a very low E_{out} .

(b) Linear Regression (applied as a classification method).

Linear Regression always took a very small amount of time and computing power to come up with its solution. Unfortunately in many experiments that solution provided a rather large E_{out} . It seems that Linear Regression has issues with messy data, where finding a clean pattern is difficult or impossible. In terms of computing resources and time it came out on top, but the large E_{out} makes it pretty unsuitable for classification purposes.

(c) The Pocket algorithm, starting from the solution given by linear regression.

Unfortunately I could not figure out how to load in the linear regression solution into w without getting an error.

(Extra) Add significant outliers to the $y = +1$ class and explain how it affects your results.

Same problem as (c). I could not figure out how to directly manipulate the data without getting errors.

Problem 2

(a)

Show that:

$$\frac{d\sigma}{d\alpha} = \sigma(\alpha)(1 - \sigma(\alpha))$$

If $\frac{d\sigma}{d\alpha} = \frac{e^\alpha}{(1+e^\alpha)^2}$ then we can prove this by the following.

$$\sigma(\alpha)(1 - \sigma(\alpha)) = \left(\frac{1}{1 + e^{-\alpha}}\right)\left(1 - \frac{1}{1 + e^{-\alpha}}\right)$$

$$1 - \frac{1}{1 + e^{-\alpha}} = \frac{e^{-\alpha}}{1 + e^{-\alpha}}$$

$$\left(\frac{1}{1 + e^{-\alpha}}\right)\left(\frac{e^{-\alpha}}{1 + e^{-\alpha}}\right) = \frac{e^{-\alpha}}{(1 + e^{-\alpha})^2}$$

(b) Derive the gradient of the log-likelihood.

The gradient of the log-likelihood is equal to our true output minus our hypothesis times x , so using the chain rule, $\nabla_w l(w) = (y - h(x))x$.

(c) Using the gradient we just found, write the update step for gradient ascent of $l(w)$.

If our update step is originally $w = w + (s^*x)$, the our update step using the gradient becomes $w = (s-h(x))^*x$.