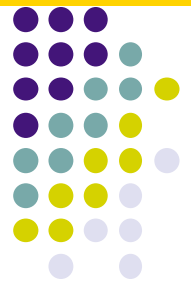


# Homework # 3 first order methods Regularized Logistic Regression



## HW #3: gradient descent for logistic regression

Consider the following optimization problem ( $l_2$  regularized logistic regression (note that there is no intercept  $\beta$ , so modify data accordingly, to add one extra dimension to  $w$  for the intercept.

$$\min_w f(w) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(w^\top x_i))) + \lambda \|w\|_2^2$$

Apply the simple proximal gradient algorithm:

$$w^{k+1} = \operatorname{argmin}_v Q_{f, \mu_k}(w^k, v)$$

$$Q_{f, \mu_k}(w, v) = f(w) + \nabla f(w)^\top (v - w) + \frac{1}{2\mu_k} \|w - v\|^2$$

In other words:



$$w^{k+1} = w^k - \mu_k \nabla f(w^k)$$

Find  $\mu_k$  at each iteration by using backtracking, until the condition below is satisfied.

$$f(w^{k+1}) \leq Q_{f, \mu_k}(w^k, w^{k+1})$$

## Homework #3: gradient descent for logistic regression



For the same  $l_2$  regularized logistic regression optimization problem

$$\min_w f(w) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(w^\top x_i))) + \lambda \|w\|_2^2$$

Apply the optimal proximal gradient algorithm:

$$w^k = u^k - \mu_k \nabla f(u^k)$$

$$t_{k+1} := (1 + \sqrt{1 + 4t_k^2})/2$$

$$u^{k+1} := w^k + \frac{t_k - 1}{t_{k+1}} [w^k - w^{k-1}]$$

Find  $\mu_k$  at each iteration by using backtracking, until the condition below is satisfied. This time make sure that  $\mu_k$  is not increasing from one iteration to another.

$$f(w^{k+1}) \leq Q_{f, \mu_k}(u^k, w^{k+1})$$

# HW#3



- Implement your algorithms in Matlab.
- Stop each algorithm when the gradient becomes reasonably small (try values  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-2}$ ) or when the progress is too small.
- Test them first on the simple 2-dimensional separable data data that you generated for the IPM code.
- Then perform classification and then testing on the digits data from homework #2.
- Show that you are obtaining “reasonable” results, by comparing the testing accuracy with that of the solution produced by Mosek.