

# Statistical and Mathematical Methods for Machine Learning (SMM)

- Homework 1: Linear Algebra;
- Homework 2: PCA and LDA;
- Homework 3: Gradient Descent and Stochastic Gradient Descent;
- Homework 4: MLE and MAP;

## Homework 1: Linear Algebra and Floating Point Arithmetic

### Linear Algebra

- Consider the random matrix. Describe the Behavior and relationship between  $K_2(A)$  and  $K_\infty(A)$ . Why their overall trend is similar? Does the definition of ill-conditioning depend on the norm used? Is there a relationship between the condition number of a matrix and the relative error  $E(x_{true}, x)$  of the computed solution?
- Consider the Vandermonde matrix. Describe the behavior and relationship between  $K_2(A)$  and  $K_\infty(A)$ . Why their overall trend is similar? Does the definition of ill-conditioning depend on the norm used? Is there a relationship between the condition number of a matrix and the relative error  $E(x_{true}, x)$  of the computed solution?
- Consider the Hilbert matrix. Describe the behavior and relationship between  $K_2(A)$  and  $K_\infty(A)$ . Why their overall trend is similar? Does the definition of ill-conditioning depend on the norm used? Is there a relationship between the condition number of a matrix and the relative error  $E(x_{true}, x)$  of the computed solution?

## Homework 2: PCA and LDA

### PCA and LDA

- Compare the ability of PCA and LDA in clustering the data. Explain the different behavior of the two methods;
- Compute the average distance of the centroid of the two methods in the training and test set. There are some differences?
- Define the classifier associated with PCA and LDA. Compare the results. Why does it happens?
- Define the classifier associated with PCA and LDA. What happens to the accuracy when  $k$  grows? Why the accuracy over the test set does not increase monotonically?

### SVD Decomposition (Dyads)

- Consider two different images. What do you observe if you compare the  $k$  rank approximation of an image  $X$  for increasing values of  $k$ ? Is there a relationship between the meaningfulness of the dyad of  $X$  for a given  $k$

and the value of the associated singular value? What do you observe if you plot the approximation error  $\|X_k - X\|_2$  compared with the plot of  $\sigma_k$ , for increasing values of  $k$ ?

## Homework 3: Optimization

### Gradient Descent (GD)

- Comparison between GD with and without backtracking (for different  $\alpha > 0$ ). What is the behavior for different functions? Explain.
- By looking the plots of  $\|\nabla f(x^k)\|_2$ , of the error  $\|x_{TRUE} - x_K\|_2$  and of  $\|x^k - x^*\|$ , compare the convergence speed for different functions and for different values of  $\alpha > 0$ , constant and chosen with backtracking procedure.
- Consider the function 1. By looking the plots of  $\|\nabla f(x^k)\|_2$ , of the error  $\|x_{TRUE} - x_K\|_2$  and of  $\|x^k - x^*\|$ , discuss the convergence by changing the starting iterate, the tolerances and the step size. Optional:\* Observe the behavior of the contour plot for the given examples.
- Consider the function 2. By looking the plots of  $\|\nabla f(x^k)\|_2$ , of the error  $\|x_{TRUE} - x_K\|_2$  and of  $\|x^k - x^*\|$ , discuss the convergence by changing the starting iterate, the tolerances and the step size. Optional:\* Observe the behavior of the contour plot for the given examples.
- Consider the function 3. By looking the plots of  $\|\nabla f(x^k)\|_2$ , of the error  $\|x_{TRUE} - x_K\|_2$  and of  $\|x^k - x^*\|$ , discuss the convergence by changing the value of  $n$  as in the homework trace, the tolerances and the step size.
- Consider the function 4. By looking the plots of  $\|\nabla f(x^k)\|_2$ , of the error  $\|x_{TRUE} - x_K\|_2$  and of  $\|x^k - x^*\|$ , discuss the convergence by changing the the value of  $n$  as in the homework trace, the tolerances and the step size.
- Consider the function 5. Discuss the point of GD with different values of  $x_0$  and different step-sizes. Observe when the convergence points the global minimum and when it stops on a local minimum or maximum.

### Stochastic Gradient Descent (SGD)

- Discuss the behavior of the logistic regression classifier varying the training set dimension ( $N_{train}$ ).
- Discuss the behavior of the logistic regression classifier varying the two considered digits.
- What are the differences at convergence of the parameters  $w^*$  when computed by GD and SGD, in particular the error of  $w^*$  against the true solution.
- Compare the accuracy of the Logistic Regression Classifier against PCA and LDA for the same considered digits (two digits only)
- *Optional:* Compare the accuracy of the Logistic Regression Classifier against PCA and LDA for three digits for the same considered digits.

## Homework 4: MLE and MAP

- What is the behavior of the trained regressor model  $f_{\theta}(x)$ , where  $\theta$  is the MLE solution under Gaussian assumptions, for increasing values of  $K$ ? Explain the plot where the training and the test error are compared for increasing values of  $K$ .
- What is the behavior of the trained regressor model  $f_{\theta}(x)$ , where  $\theta$  is the MAP solution under Gaussian assumptions, for increasing values of  $K$  and fixed  $\lambda$ ? Explain the plot where the training and the test error are compared for increasing values of  $K$ .
- What is the behavior of the trained regressor model  $f_{\theta}(x)$ , where  $\theta$  is the MAP solution under Gaussian assumptions, for fixed value  $K$  lower and/or greater than the true  $K$  and different  $\lambda$ ?
- Comment the difference in relative error between the MLE and MAP solutions, for given  $\lambda > 0$  and increasing  $K$ . What happens when  $N$  increases, if everything else stays the same? What are the differences between the solution computed via GD, SGD and Normal Equations? Does the relative error between the computed weights and the true weights relates with the accuracy of the computed model? Explain.
- *Optionals:* The previous with Poisson noise in place of gaussian noise.