Lonari, Nikit
Homework 1
Machine Learning

For **2a,**

the distance between each new data point and the training data point is calculated using cartesian distance and stored in matrix named distance, as follows

Editor – knn_algorithm.m

distance

14x4 double

	1	2	3	4
1	9.7980	18.1108	14.0712	29.8831
2	50.4777	30	29.2233	14.1774
3	16.0312	35.1141	35.7071	50.3587
4	13.4536	10.4403	8.6603	23.3238
5	28.2489	9.6954	8.9443	9.1104
6	43.7493	22.6716	22.4499	7.6811
7	15	11.1803	6.7082	21.9545
8	9.1104	29.9833	29.7489	44.6990
9	7.6811	16.0935	14.4568	29.2233
10	33.6601	15	12.2066	5.4772
11	20.6155	10.0499	2.2361	16.7929
12	31.4006	12.2474	10.8628	5.7446
13	4.6904	26.2679	25.0200	40.3609
14	20.7123	4.1231	5.3852	16.4317
15				

then, sort each column in distance matrix such that there indices remains the same(primitive); which I have stored the sorted data in order matrix and their id's in indices.
User then inputs how many K- nearest neighbours to be calculated (1,3,5).
Then I took the K-nearest neighbours from order matrix and their respective id's to retrieve the value whether Women/Men from training data set.
Simultaneously, a counter for Women/Men is updated simultaneously for each step of KNN.

Predictions, are been given as below,

for K=1		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	W
	(180,85,29)	M
K=3		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	W
	(180,85,29)	M
K=5		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	M
	(180,85,29)	M

For **2c**,
distance is calculated without age values,

Editor - knn_algorithm_for_2c.m

distance | indices

14x4 double

	1	2	3	4
1	8.9443	18.1108	13.9284	29.7321
2	50.4777	29.7321	29.1548	14.1421
3	14.4222	34.9857	35.3553	50
4	13.4164	10	8.6023	23.3238
5	28.1780	7.6158	8	8.6023
6	43.5660	22.6716	22.3607	7.0711
7	15	10.4403	6.4031	21.9317
8	8.6023	29.9666	29.7321	44.6542
9	7.6158	15.2971	14.1421	29.1548
10	33.6006	14.8661	12.2066	5.3852
11	20.6155	9.2195	1	16.7631
12	31.3847	11.1803	10.4403	5.3852
13	3.6056	26.2488	25	40.3113
14	20.6155	3.6056	5.3852	16.4012

When age attribute is removed we get nearly all same prediction up till K=5;
The table below shows prediction after removing the age attribute.

for K=1		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	W
	(180,85,29)	M
K=3		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	W
	(180,85,29)	M
K=5		
	(162,53,28)	W
	(168,75,32)	M
	(175,70,30)	M
	(180,85,29)	M

Unfortunately, age in the given dataset is such an attribute which fails to provide additive information when compared with the information from height and weight of an individual to be a women or men. To conclude, we can easily have the same prediction using height and weight because the age values are not sparsely distributed in xyz co-ordinate axis.