CSE 6363 - Machine Learning

Fall 2022

Due Date: October 9, 2022, 11:59 PM

Data Set

Use the dataset given at the bottom of this file.

K Nearest Neighbor

Q1 Consider the problem where we want to predict the gender of a person from a set of input parameters, namely height, weight, and age.

a) Using **Cartesian distance**, **Manhattan distance and Minkowski distance of order 3** as the similarity measurements show the results of the gender prediction for the Evaluation data that is listed below generated training data for values of K of **1, 3, and 7**. **Include the intermediate steps** (i.e., distance calculation, neighbor selection, and prediction).
b) Implement the KNN algorithm for this problem**.** Your implementation should work **with different training data sets** as well **as different values of K and allow to input a data point for the predictio**n.
c) To evaluate the performance of the KNN algorithm (**using Euclidean distance metric**), implement a **leave-one-out evaluation** routine for your algorithm. In leave-one-out validation, we repeatedly evaluate the algorithm by removing one data point from the training set, training the algorithm on the remaining data set and then testing it on the point we removed to see if the label matches or not. Repeating this for each of the data points gives us an estimate as to the **percentage of erroneous predictions** the algorithm makes and thus a measure of the accuracy of the algorithm for the given data. Apply your leave-one-out validation with your KNN algorithm to the dataset for Question **1 c)** for values for K of 1, 3, 5, 7, 9, and 11 and report the results. For which value of K do you get the best performance?
d) Repeat the prediction and validation you performed in Question **1 c)** using KNN when the age data is removed (i.e. when only the height and weight features are used as part of the distance calculation in the KNN algorithm). Report the results and compare the performance without the age attribute with the ones from Question **1 c).** Discuss the results. What do the results tell you about the data?

Gaussian Na¨ive Bayes Classification

Q2. Using the data from Problem 2, build a Gaussian Na¨ive Bayes classifier for this problem. For this you have to learn Gaussian distribution parameters for each input data feature, i.e. for p(height|W), p(height|M), p(weight|W), p(weight|M), p(age|W), p(age|M).

a) Learn/derive the parameters for the Gaussian Na¨ive Bayes Classifier for the data from Question 2 a) and apply them to the same target as in problem **1a)**.
b) Implement the Gaussian Na¨ive Bayes Classifier for this problem.
c) Repeat the experiment in **part 1 c) and 1 d)** with the Gaussian Na¨ive Bayes Classifier. Discuss the results, in particular with respect to the performance difference between using all features and using only height and weight.
d) Same as 1d but with Naïve Bayes.
e) Compare the results of the two classifiers (i.e., the results form 1 c) and 1d) with the ones from 2 c) 2d) and discuss reasons why one might perform better than the other.

Continued on next page →

Some rules to follow:

1. **Handwrite, sign, and date (with date of submission) a copy of the Honor Code (shown below) and share the image as part of your project; a handwritten, signed, and dated (with the date of submission) copy of the Honor Code must be included with every project and exam submission. (Failing to include will cost 20 points)**
2. **Students are required to NOT share their solutions to the project even after the semester is over or even after graduation. However, they can show their projects during their interviews. They are also required to not discuss the solution with others or use anyone else's solution. Any violation of the policy will result in a 0 for this project for all students concerned.**

HONOR CODE

I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.

I promise that I will submit only work that I personally create or that I contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code

I will not participate in any form of cheating/sharing the questions/solutions.

Question 1 a) and 2 a) Training Data:

((1.6530190426733, 72.871146648479, 24), W)
((1.6471384909498, 72.612785314988, 34), W)
((1.6472055785348, 73.53968351051, 33), M)
((1.7323008914951, 76.067870338779, 30), M)
((1.6750702657911, 81.05582111533, 30), M)
((1.5780970716644, 64.926084680188, 30), W)
((1.6587629355524, 69.38092449041, 30), M)
((1.6763295980234, 77.062295990149, 31), M)
((1.7187224085504, 62.112923317057, 37), W)
((1.5202218226439, 66.151444019603, 27), W)
((1.5552689261884, 66.076386143769, 31), W)
((1.6969333189258, 77.45386244568, 34), M)
((1.6887980792886, 76.489640732464, 37), M)
((1.5213552893624, 63.952944947832, 35), W)

Question 1 a) and 2 a) Test Data:

(1.62065758929, 59.376557437583, 32)
(1.7793983848363, 72.071775670801, 36)
(1.7004576585974, 66.267508112786, 31)
(1.6591086215159, 61.751621901787, 29)

Question 1 c), 1 d), 2 c), 2 d) Program Data

((1.5963600450124, 75.717194178189, 23), W)
((1.6990610819676, 83.477307503684, 25), M)
((1.5052092436, 74.642420817737, 21), W)
((1.5738635789008, 78.562465284603, 30), M)
((1.796178772769, 74.566117057707, 29), M)
((1.6274618774347, 82.250591567161, 21), W)
((1.6396843250708, 71.37567170848, 20), W)
((1.538505823668, 77.418902097029, 32), W)
((1.6488692005889, 76.333044488477, 26), W)
((1.7233804613095, 85.812112126306, 27), M)
((1.7389100516771, 76.424421782215, 24), W)
((1.5775696242624, 77.201404139171, 29), W)

((1.7359417237856, 77.004988515324, 20), M)
((1.5510482441354, 72.950756316157, 24), W)
((1.5765653263667, 74.750113664457, 34), W)
((1.4916026885377, 65.880438515643, 28), W)
((1.6755053770068, 78.901754249459, 22), M)
((1.4805881225567, 69.652364469244, 30), W)
((1.6343943760912, 73.998278712613, 30), W)
((1.6338449829543, 79.216500811112, 27), W)
((1.5014451222259, 66.917339299419, 27), W)
((1.8575887178701, 79.942454850988, 28), M)
((1.6805940669394, 78.213519314007, 27), W)
((1.6888905106948, 83.031099742808, 20), M)
((1.7055120272359, 84.233282531303, 18), M)
((1.5681965896812, 74.753880204215, 22), W)
((1.6857758389206, 84.014217544019, 25), W)
((1.7767370337678, 75.709336556562, 27), M)
((1.6760125952287, 74.034126149139, 28), M)
((1.5999112612548, 72.040030344184, 27), M)
((1.6770845322305, 76.149431872551, 25), M)
((1.7596128136991, 87.366395298795, 29), M)
((1.5344541456027, 73.832214971449, 22), W)
((1.5992629534387, 82.4806916967, 34), W)
((1.6714162787917, 67.986534194515, 29), W)
((1.7070831676329, 78.269583353177, 25), M)
((1.5691295338456, 81.09431696972, 27), M)
((1.7767893419281, 76.910413184648, 30), M)
((1.5448153215763, 76.888087599642, 32), W)
((1.5452842691008, 69.761889289463, 30), W)
((1.6469991919639, 82.289126983444, 18), W)
((1.6353732734723, 77.829257585654, 19), W)
((1.7175342426502, 85.002276406574, 26), M)
((1.6163551692382, 77.247935733799, 21), M)
((1.6876845881843, 85.616829192322, 27), M)
((1.5472705508274, 64.474350365634, 23), W)
((1.558229415357, 80.382011318379, 21), W)
((1.6242189230632, 69.567339939973, 28), W)
((1.8215645865237, 78.163631826626, 22), W)
((1.6984142478298, 69.884030497097, 26), M)
((1.6468551415123, 82.666468220128, 29), M)
((1.5727791290292, 75.545348033094, 24), M)
((1.8086593470477, 78.093913654921, 27), M)
((1.613966988578, 76.083586505149, 23), W)
((1.6603990297076, 70.539053122611, 24), M)
((1.6737443242383, 66.042005829182, 28), W)
((1.6824912337281, 81.061984274536, 29), M)
((1.5301691510101, 77.26547501308, 22), M)
((1.7392340943261, 92.752488433153, 24), M)
((1.6427105169884, 83.322790265985, 30), M)
((1.5889040551166, 74.848224733663, 25), W)
((1.5051718284868, 80.078271153645, 31), W)
((1.729420786579, 81.936423109142, 26), M)
((1.7352568354092, 85.497712687992, 19), M)
((1.5056950011245, 73.726557750383, 24), W)
((1.772404089054, 75.534265951718, 30), M)
((1.5212346939173, 74.355845722315, 29), W)
((1.8184515409355, 85.705767969326, 25), M)
((1.7307897479464, 84.277029918205, 28), W)
((1.6372690389158, 72.289040612489, 27), M)
((1.6856953072545, 70.406532419182, 28), W)
((1.832494802635, 81.627925524191, 27), M)
((1.5061197864796, 85.886760677468, 31), W)
((1.5970906671458, 71.755566818152, 27), W)
((1.6780459059283, 78.900587239209, 25), W)
((1.6356901170146, 84.066566323977, 21), W)
((1.6085494116591, 70.950456539016, 30), M)

((1.5873479102442, 77.558144903338, 25), M)
((1.7542078120838, 75.3117550236, 26), M)
((1.642417315747, 67.97377818999, 31), W)
((1.5744266340913, 81.767568318602, 23), M)
((1.8470601407979, 68.606183538532, 30), W)
((1.7119387468283, 80.560922353487, 27), W)
((1.6169930563306, 75.538611935125, 27), M)
((1.6355653058986, 78.49626023408, 24), M)
((1.6035395957618, 79.226052358485, 33), M)
((1.662787957279, 76.865925681154, 25), M)
((1.5889291137091, 76.548543553914, 28), W)
((1.9058127964477, 82.56539915922, 25), M)
((1.694633493614, 62.870480634419, 21), W)
((1.7635692396034, 82.479783004684, 27), M)
((1.6645292231449, 75.838104636904, 29), W)
((1.7201968406129, 81.134689293557, 24), W)
((1.5775563651749, 65.920103519266, 24), W)
((1.6521294216004, 83.312640709417, 28), M)
((1.5597501915973, 76.475667826389, 30), W)
((1.7847561120027, 83.363676219109, 29), M)
((1.6765690500715, 73.98959022721, 23), M)
((1.6749260607992, 73.687015573315, 27), W)
((1.58582362825, 71.713707691505, 28), M)
((1.5893375739649, 74.248033504548, 27), W)
((1.6084440045081, 71.126430164213, 27), W)
((1.6048804804343, 82.049319162211, 26), W)
((1.5774196609804, 70.878214496062, 24), W)
((1.6799586185525, 75.649534976838, 29), W)
((1.7315642636281, 92.12183674186, 29), M)
((1.5563282000349, 69.312673560451, 32), W)
((1.7784349641893, 83.464562543, 26), M)
((1.7270244609765, 76.599791001341, 22), W)
((1.6372540837311, 74.746741127229, 30), W)
((1.582550559056, 73.440027907722, 23), W)
((1.722864383186, 79.37821152354, 20), W)
((1.5247544081009, 70.601290492141, 27), W)
((1.580858666774, 70.146982323579, 24), W)
((1.703343390074, 90.153276095421, 22), W)
((1.5339948635367, 59.675627532338, 25), W)
((1.8095306490733, 86.001187990639, 20), M)
((1.7454786971676, 85.212429336602, 22), M)
((1.6343303342105, 85.46378358014, 32), M)
((1.5983479173071, 79.323905480504, 27), W)