

DSC 433/533 – Class 13 – Nearest Neighbors example

Task: classify a credit card holder as a likely responder to an offer (or not) and also predict his/her average monthly balance (in \$) based on:

- Gender (M or F);
- age (in years);
- salary (in \$ thousands).

Training data for 4 credit card holders:

	Gender	Age	Salary	Response	Balance
Anne	F	25	50	No	300
Bob	M	22	55	No	600
Carrie	F	23	55	Yes	400
Doug	M	26	40	Yes	500

Eric: 24 year-old male earning \$45k

Distances	Gender	Age	Std. Age	Salary	Std. Salary	Manhattan	Euclidean
Anne	1	1	0.5	5	0.5	2.0 (2=)	1.22 (2)
Bob	0	2	1.0	10	1.0	2.0 (2=)	1.41 (3)
Carrie	1	1	0.5	10	1.0	2.5 (4)	1.50 (4)
Doug	0	2	1.0	5	0.5	1.5 (1)	1.12 (1)
Maximum		2		10			

Gender distances are 0 for a match, 1 for a mismatch.

Standardized distances for quantitative variables are the absolute difference divided by the maximum absolute difference.

Manhattan distance is $G + A + S$.

Euclidean distance is $\sqrt{G^2 + A^2 + S^2}$.

Neighbors (k)	Manhattan classification	Manhattan prediction	Euclidean classification	Euclidean prediction
1	1 Y, so classify Y	\$500	1 Y, so classify Y	\$500
2	1 Y, 2 N, so classify N	$(500+300+600)/3=\$467$	1 Y, 1 N, so classify ?	$(500+300)/2=\$400$
3	1 Y, 2 N, so classify N	$(500+300+600)/3=\$467$	1 Y, 2 N, so classify N	$(500+300+600)/3=\$467$

Fiona: 22 year-old female earning \$50k

Distances	Gender	Age	Std. Age	Salary	Std. Salary	Manhattan	Euclidean
Anne	0	3	0.75	0	0.0	0.75 (1=)	0.75 (2)
Bob	1	0	0.00	5	0.5	1.50 (3)	1.12 (3)
Carrie	0	1	0.25	5	0.5	0.75 (1=)	0.56 (1)
Doug	1	4	1.00	10	1.0	3.00 (4)	1.73 (4)
Maximum		4		10			

Neighbors (k)	Manhattan classification	Manhattan prediction	Euclidean classification	Euclidean prediction
1	1 Y, 1 N, so classify ?	$(400+300)/2=\$350$	1 Y, so classify Y	\$400
2	1 Y, 1 N, so classify ?	$(400+300)/2=\$350$	1 Y, 1 N, so classify ?	$(400+300)/2=\$350$
3	1 Y, 2 N, so classify N	$(400+300+600)/3=\$433$	1 Y, 2 N, so classify N	$(400+300+600)/3=\$433$