Steps to Follow for the k-NN task:

(I) Do the data conversion:

(a) for non-numeric attributes, if the similarity matrix is not provided, convert them (eg: high=2, med=1, low=0)

(b) Normalize all the numeric attributes to fall between 0 and 1 using the formula on slide 33. (slide 33 also has an example)

(II) for each vector in Dtest do the following:

(1) Use the formula on slide 35 to find the similarity between the test vector to all training data vectors

(1a) Notice that the formula on slide 35 takes the difference of the attribute values (between the test and train) in the denominator. You can take the difference for numeric attributes but for non-numeric attributes, use the corresponding similarity value from the similarity matrix.
For example, if attribute 3 is nonnumeric in the train data vector ($a_1\ a_2\ ...\ a_n$) and test data vector ($b_1 b_2 ..... b_n$)

$$\frac{1}{\sqrt{\left[\ (a_1-b_1)^2 + \ (a_2-b_2)^2 + (1-\text{sim}(a_3,b_3)) \ + (a_4-b_4)^2 + ... + (a_n-b_n)^2\ \right]}}$$

(2) sort and find the top 3 closest training vectors using

$$\text{Value}_{obj}(y) = \text{argmax}_{Ci}\ [\ \Sigma_{x_j \in kNN(y)}\ \ \text{sim}(x_j,y) * \delta\ (\ \ \text{class}(x_j)\ ,\ C_i\ \ \ )\ ]$$

where,

$\delta\ (\ \ \text{class}(x_j)\ ,\ C_i\ \ \ ) = 1$ if, $\text{class}(x_j) = C_i$

$\delta\ (\ \ \text{class}(xj)\ ,\ Ci\ \ \ ) = 0$ if, $\text{class}(xj) \neq Ci$

contd...

For example,

(a) After you sorted the scores, say the training vectors x_1, x_{15}, x_{20} were found to be the top 3-NN training data points close to the test vector (y)

Now say, you got the following similarity scores (inverse Euclidean distance given in the previous page) for the top 3 points:

$sim(x_1,y)=3.2$, $sim(x_{15},y)=1.2$, $sim(x_{20},y)=3.7$

(b) Say, in the training data, the class of x_1 is C3 , class of x_{15} is C4, and the class of x_{20} is C3.

So, now you know that the test vector is either C3 or C4.

You have to predict whether the class of y is C3 or C4 from just x_1, x_{15} and x_{20}

(d) prediction: argmax (score for C3(=3.2+3.7) , score for C4 (=1.2))

= argmax(score for C3(=6.9) , score for C4 (=1.2))

hence, class (or label) of y is C3, as 6.9>1.2

and the score is: 6.9

Note:Meaning of argmax

argmax stands for the argument of the maximum, i.e., the set of points for which the value of the given expression attains its maximum value