8.

(a)

According to the program result
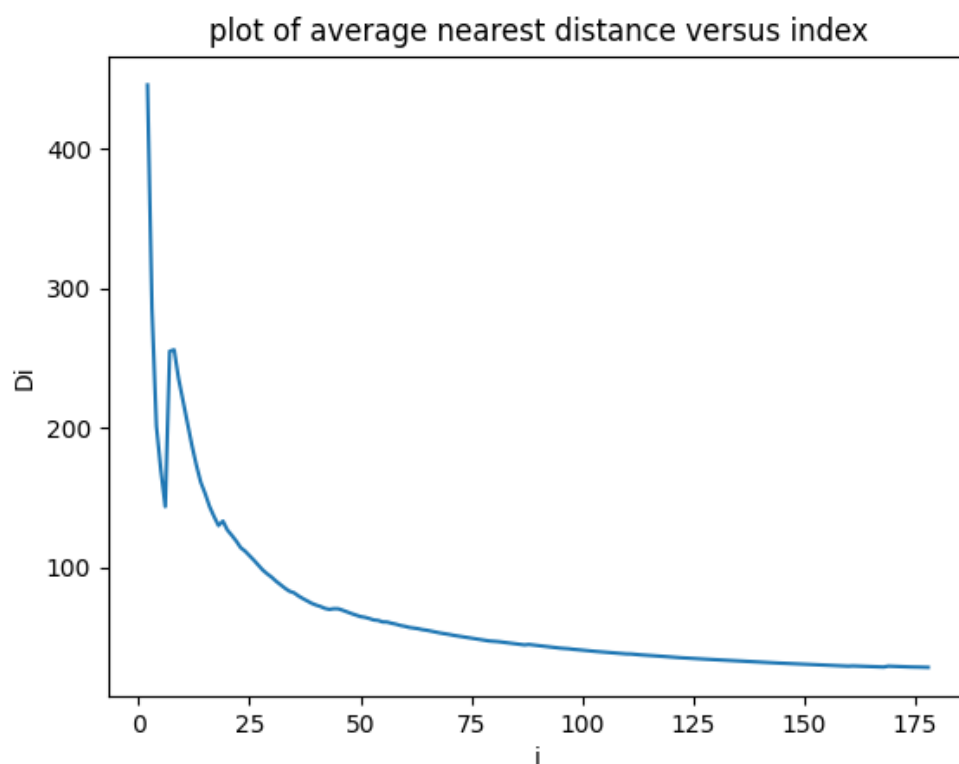
$$sum(e) = 48$$

(b)

The code is shown above

The overall trend is that the average distance to the nearest neighbor

seems to decline

Apparently, as there more and more points, it is much easier to reach

one , resulting in that the average distance declines

plot of average nearest distance versus index

(c)

The code is shown above

The overall trend is that the fraction of errors seems to decline

Apparent, as there are more and more points, the stochasticity

declines, that's why it is easier to find the "correct" neighbor.

plot of fraction of errors versus index

(d)

Apparently, different features should have different weight, also, they might be measured in different scale(unit), e.g. kg, g.
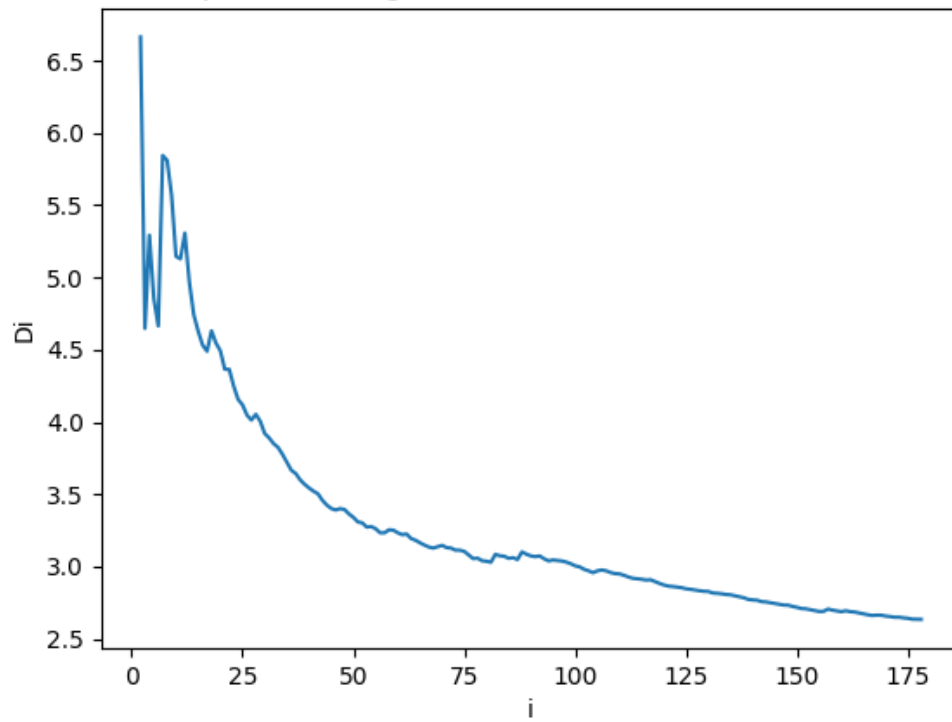
By slightly decrease the weight of the 5th feature(Magnesium) and the 13th feature(Proline), since they are obviously larger than other feature, the good result come out:

$$sum(e) = 17$$

Below is the fix in Euclidean distance

```python
def eucliddist(a,b):
    """
    compute the euclidean distance between two wines
    :param a:
    :param b:
    :return:
    """
    d2=0
    w=[1,1,1,1,0.1, 1,1,1,1,1, 1,1,0.01]
    for i in range(len(a)):
        d2+=((a[i]-b[i])*w[i])**2
    return math.sqrt(d2)
```

plot of average nearest distance versus index

plot of fraction of errors versus index