

物件導向程式設計-跨領域學習

歐式距離在人工智慧(AI)領域物件辨識之運用

【1. 作業說明】

1.不可將程式碼給別組參考，類似都不行！！

2.時間：

3.佔總成績 * %。

【2. 數學原理說明】

在人工智慧(AI)領域中，分類(classification)是一項重要的技術，我們的學過的歐式距離(Euclidean distance)通常就是一項重要的技術。一般，常用來描述一個物件的特徵(feature)或是統計上所稱的屬性(attribute)，通常高中學到 2 或 3 個，會視為空間中 2 維或 3 維的點。

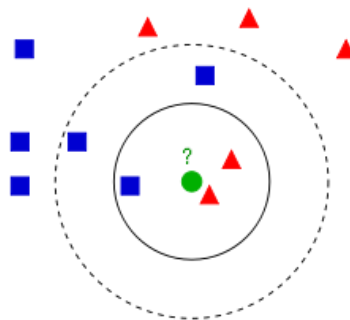
到了大學，應該將一個點一般化為一個 n 維的點 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 或向量 $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ，表示有 n 個特徵或是 n 個測量值或是 n 個屬性值，可以想像成一個人考 n 科目會得到 n 科的成績，並用數學進行記錄。因此，任兩點 \mathbf{x}, \mathbf{y} 的歐式距離就是

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

電腦或是機器通常是利用相似性測量(similarity measurement)來協助人類判定類別，距離越近表示兩者在空間中越靠近，也就越相似，因此就可以利用歐式距離來設計辨識器(classifier)。使用歐式距離的辨識器中，實用且常見的辨識器是 K 最近鄰辨識器(K-nearest neighbors[KNNs])。KNN 通常較適用於連續變數。

【3. 程式說明】

a. KNN classifier

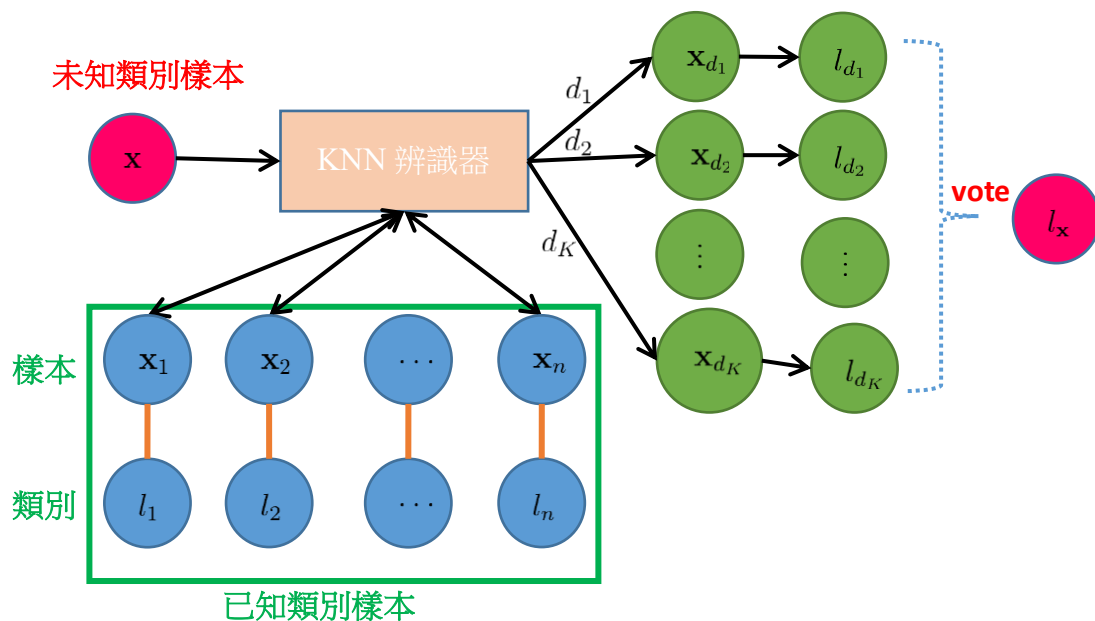


實線: K=3、虛線 K=5

俗語說「近朱者赤，近墨者黑」，「有相同羽毛的鳥會聚在一起(物以類聚)」，因此，對於一個不知道是哪一類的樣本 \mathbf{x} (上圖綠色圓圈)，怎麼知道它是哪一類？就看離它最近的 K 個已知類別的樣本，然後進行投票來作為 \mathbf{x} 的判定類別。因此利用算出倆倆距離後找最靠近的自己的 K 個已知樣本的類別進行投票，採少數

服從多數的投票策略 (majority vote)，例如 $K=3$ 就是找離 x 最近的 3 個點，假如是 a, b, c ，他們的類別分別是 1, 2, 2，則會將 2 指定給 x 作為其類別。若 x 的原始類別已知 (想像成考試真的答案)，在假設未知的情形下丟入機器產生的判定類別 (電腦給的作答)，就可以比對正確與否 (是否答對)，來知道機器的學習好壞，最常使用 **辨識正確率(classification accuracy)** 來做為量化指標(index)，以知道「某資料集使用 KNN 辨識器的效率如何」。

假設已知類別樣本之資料集 $X = \{x_1, x_2, \dots, x_n\}$ ，內含 n 個樣本點，其對應之類別 $L = \{l_1, l_2, \dots, l_n\}$ ，其中 $l_i \in 1, 2, \dots, M$ 。若未知類別樣本點 x ，因此，將 x 丟入 KNN 辨識器與 $X = \{x_1, x_2, \dots, x_n\}$ 中所有點計算歐式距離 $D = \{d_1, d_2, \dots, d_n\}$ ，並依據 K 個最近鄰的類別



b. 架構-視撰寫架構，函式越多，main()越短

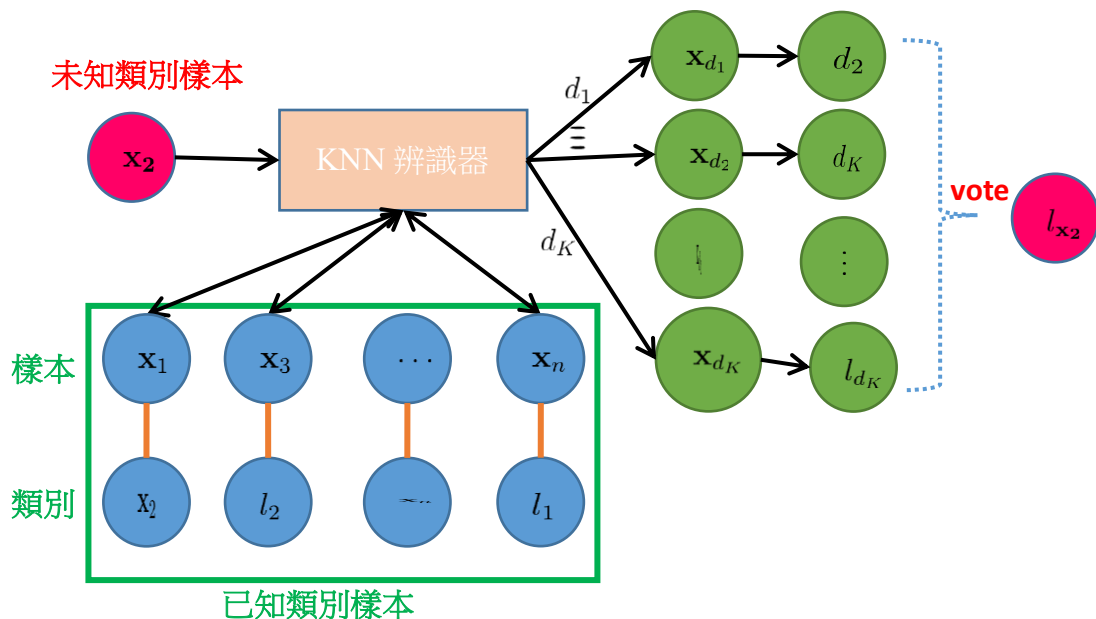
自由創作，但無使用類別概念!(分數 最高*分)

自由創作，用到類別和物件的概念!(最高*分)

自由創作，用到 2 個類別和物件的概念!(最高*分)

c. 評估資料在 KNN 機器的效能-Leave-one-out Classification (LOOC)架構-辨識正確率(classification accuracy)的計算方式

Iris 有 150 個樣本，「每一個樣本輪流當作類別未知的樣本，其他 149 個樣本當成已知樣本」。換句話說，每一個樣本有一次當未知類別樣本，149 次當成已知類別樣本。



d. 程式細節

Step 1. 讀取資料集 iris.txt 檔成資料矩陣

Step 2. 使用 LOOC 架構利用 K-Nearest Neighbors(KNNs) 辨識器 (

<https://zh.wikipedia.org/wiki/K%E8%BF%91%E9%82%BB%E7%AE%97%E6%B3%95>) 來對每一個已知類別樣本點 x 進行分類 \hat{l}_x 。

(從步驟 2 算出之矩陣中針對每個 row 找距離最近 K 個已知樣本，並根據 index 取得其類別，在根據類別中之多數，將之指定給未知樣本)

Step 3. 將所有資料中的樣本的判定類別與其真實類別 l_{d_2} 比對後，計算正確判定的比率，稱辨識率 acc。本專題針對選定 $K=1$ 或 $K=3$ 或 $K=5$ 進行 acc 計算，三個不同 K 值計算出之辨識率分別為 acc1, acc2 和 acc3。

【4. 資料集說明-安德森鳶尾花卉資料集】

安德森鳶尾花卉資料集 (英文: Anderson's Iris data set)，也稱鳶尾花卉資料集 (英文: Iris flower data set) 或費雪鳶尾花卉資料集 (英文: Fisher's Iris data set)，是一類多重變量分析的資料集。它最初是埃德加·安德森從加拿大加斯帕半島上的鳶尾屬花朵中提取的形態學變異數據，後由羅納德·費雪作為判別分析的一個例子，運用到統計學中。其數據集包含了 150 個樣本，都屬於鳶尾屬下的三個亞屬，分別是山鳶尾、變色鳶尾和維吉尼亞鳶尾 (Iris setosa, Iris versicolor and Iris virginica)。四個特徵被用作樣本的定量分析，它們分別是花萼 (Sepal) 和花瓣 (Petal) 的長度和寬度。



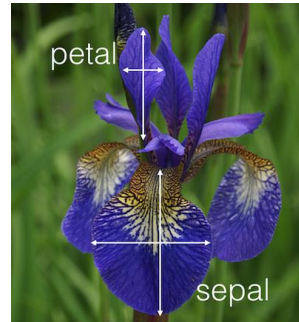
山鳶尾



變色鳶尾



維吉尼亞鳶尾



<i>sepal length</i>	<i>sepal width</i>	<i>petal length</i>	<i>petal width</i>	Label	類別名稱
5.1	3.5	1.4	0.2	1	<i>setosa</i>
4.9	3	1.4	0.2	1	<i>setosa</i>
4.7	3.2	1.3	0.2	1	<i>setosa</i>
4.6	3.1	1.5	0.2	1	<i>setosa</i>
5	3.6	1.4	0.2	1	<i>setosa</i>
5.4	3.9	1.7	0.4	1	<i>setosa</i>
4.6	3.4	1.4	0.3	1	<i>setosa</i>
5	3.4	1.5	0.2	1	<i>setosa</i>
4.4	2.9	1.4	0.2	1	<i>setosa</i>
4.9	3.1	1.5	0.1	1	<i>setosa</i>
5.4	3.7	1.5	0.2	1	<i>setosa</i>
4.8	3.4	1.6	0.2	1	<i>setosa</i>
4.8	3	1.4	0.1	1	<i>setosa</i>
4.3	3	1.1	0.1	1	<i>setosa</i>
5.8	4	1.2	0.2	1	<i>setosa</i>
5.7	4.4	1.5	0.4	1	<i>setosa</i>
5.4	3.9	1.3	0.4	1	<i>setosa</i>
5.1	3.5	1.4	0.3	1	<i>setosa</i>
5.7	3.8	1.7	0.3	1	<i>setosa</i>
5.1	3.8	1.5	0.3	1	<i>setosa</i>

5.4	3.4	1.7	0.2	1	<i>setosa</i>
5.1	3.7	1.5	0.4	1	<i>setosa</i>
4.6	3.6	1	0.2	1	<i>setosa</i>
5.1	3.3	1.7	0.5	1	<i>setosa</i>
4.8	3.4	1.9	0.2	1	<i>setosa</i>
5	3	1.6	0.2	1	<i>setosa</i>
5	3.4	1.6	0.4	1	<i>setosa</i>
5.2	3.5	1.5	0.2	1	<i>setosa</i>
5.2	3.4	1.4	0.2	1	<i>setosa</i>
4.7	3.2	1.6	0.2	1	<i>setosa</i>
4.8	3.1	1.6	0.2	1	<i>setosa</i>
5.4	3.4	1.5	0.4	1	<i>setosa</i>
5.2	4.1	1.5	0.1	1	<i>setosa</i>
5.5	4.2	1.4	0.2	1	<i>setosa</i>
4.9	3.1	1.5	0.2	1	<i>setosa</i>
5	3.2	1.2	0.2	1	<i>setosa</i>
5.5	3.5	1.3	0.2	1	<i>setosa</i>
4.9	3.6	1.4	0.1	1	<i>setosa</i>
4.4	3	1.3	0.2	1	<i>setosa</i>
5.1	3.4	1.5	0.2	1	<i>setosa</i>
5	3.5	1.3	0.3	1	<i>setosa</i>
4.5	2.3	1.3	0.3	1	<i>setosa</i>
4.4	3.2	1.3	0.2	1	<i>setosa</i>
5	3.5	1.6	0.6	1	<i>setosa</i>
5.1	3.8	1.9	0.4	1	<i>setosa</i>
4.8	3	1.4	0.3	1	<i>setosa</i>
5.1	3.8	1.6	0.2	1	<i>setosa</i>
4.6	3.2	1.4	0.2	1	<i>setosa</i>
5.3	3.7	1.5	0.2	1	<i>setosa</i>
5	3.3	1.4	0.2	1	<i>setosa</i>
7	3.2	4.7	1.4	2	<i>versicolor</i>
6.4	3.2	4.5	1.5	2	<i>versicolor</i>
6.9	3.1	4.9	1.5	2	<i>versicolor</i>
5.5	2.3	4	1.3	2	<i>versicolor</i>
6.5	2.8	4.6	1.5	2	<i>versicolor</i>
5.7	2.8	4.5	1.3	2	<i>versicolor</i>

6.3	3.3	4.7	1.6	2	<i>versicolor</i>
4.9	2.4	3.3	1	2	<i>versicolor</i>
6.6	2.9	4.6	1.3	2	<i>versicolor</i>
5.2	2.7	3.9	1.4	2	<i>versicolor</i>
5	2	3.5	1	2	<i>versicolor</i>
5.9	3	4.2	1.5	2	<i>versicolor</i>
6	2.2	4	1	2	<i>versicolor</i>
6.1	2.9	4.7	1.4	2	<i>versicolor</i>
5.6	2.9	3.6	1.3	2	<i>versicolor</i>
6.7	3.1	4.4	1.4	2	<i>versicolor</i>
5.6	3	4.5	1.5	2	<i>versicolor</i>
5.8	2.7	4.1	1	2	<i>versicolor</i>
6.2	2.2	4.5	1.5	2	<i>versicolor</i>
5.6	2.5	3.9	1.1	2	<i>versicolor</i>
5.9	3.2	4.8	1.8	2	<i>versicolor</i>
6.1	2.8	4	1.3	2	<i>versicolor</i>
6.3	2.5	4.9	1.5	2	<i>versicolor</i>
6.1	2.8	4.7	1.2	2	<i>versicolor</i>
6.4	2.9	4.3	1.3	2	<i>versicolor</i>
6.6	3	4.4	1.4	2	<i>versicolor</i>
6.8	2.8	4.8	1.4	2	<i>versicolor</i>
6.7	3	5	1.7	2	<i>versicolor</i>
6	2.9	4.5	1.5	2	<i>versicolor</i>
5.7	2.6	3.5	1	2	<i>versicolor</i>
5.5	2.4	3.8	1.1	2	<i>versicolor</i>
5.5	2.4	3.7	1	2	<i>versicolor</i>
5.8	2.7	3.9	1.2	2	<i>versicolor</i>
6	2.7	5.1	1.6	2	<i>versicolor</i>
5.4	3	4.5	1.5	2	<i>versicolor</i>
6	3.4	4.5	1.6	2	<i>versicolor</i>
6.7	3.1	4.7	1.5	2	<i>versicolor</i>
6.3	2.3	4.4	1.3	2	<i>versicolor</i>
5.6	3	4.1	1.3	2	<i>versicolor</i>
5.5	2.5	4	1.3	2	<i>versicolor</i>
5.5	2.6	4.4	1.2	2	<i>versicolor</i>
6.1	3	4.6	1.4	2	<i>versicolor</i>

5.8	2.6	4	1.2	2	<i>versicolor</i>
5	2.3	3.3	1	2	<i>versicolor</i>
5.6	2.7	4.2	1.3	2	<i>versicolor</i>
5.7	3	4.2	1.2	2	<i>versicolor</i>
5.7	2.9	4.2	1.3	2	<i>versicolor</i>
6.2	2.9	4.3	1.3	2	<i>versicolor</i>
5.1	2.5	3	1.1	2	<i>versicolor</i>
5.7	2.8	4.1	1.3	2	<i>versicolor</i>
6.3	3.3	6	2.5	3	<i>virginica</i>
5.8	2.7	5.1	1.9	3	<i>virginica</i>
7.1	3	5.9	2.1	3	<i>virginica</i>
6.3	2.9	5.6	1.8	3	<i>virginica</i>
6.5	3	5.8	2.2	3	<i>virginica</i>
7.6	3	6.6	2.1	3	<i>virginica</i>
4.9	2.5	4.5	1.7	3	<i>virginica</i>
7.3	2.9	6.3	1.8	3	<i>virginica</i>
6.7	2.5	5.8	1.8	3	<i>virginica</i>
7.2	3.6	6.1	2.5	3	<i>virginica</i>
6.5	3.2	5.1	2	3	<i>virginica</i>
6.4	2.7	5.3	1.9	3	<i>virginica</i>
6.8	3	5.5	2.1	3	<i>virginica</i>
5.7	2.5	5	2	3	<i>virginica</i>
5.8	2.8	5.1	2.4	3	<i>virginica</i>
6.4	3.2	5.3	2.3	3	<i>virginica</i>
6.5	3	5.5	1.8	3	<i>virginica</i>
7.7	3.8	6.7	2.2	3	<i>virginica</i>
7.7	2.6	6.9	2.3	3	<i>virginica</i>
6	2.2	5	1.5	3	<i>virginica</i>
6.9	3.2	5.7	2.3	3	<i>virginica</i>
5.6	2.8	4.9	2	3	<i>virginica</i>
7.7	2.8	6.7	2	3	<i>virginica</i>
6.3	2.7	4.9	1.8	3	<i>virginica</i>
6.7	3.3	5.7	2.1	3	<i>virginica</i>
7.2	3.2	6	1.8	3	<i>virginica</i>
6.2	2.8	4.8	1.8	3	<i>virginica</i>
6.1	3	4.9	1.8	3	<i>virginica</i>

6.4	2.8	5.6	2.1	3	<i>virginica</i>
7.2	3	5.8	1.6	3	<i>virginica</i>
7.4	2.8	6.1	1.9	3	<i>virginica</i>
7.9	3.8	6.4	2	3	<i>virginica</i>
6.4	2.8	5.6	2.2	3	<i>virginica</i>
6.3	2.8	5.1	1.5	3	<i>virginica</i>
6.1	2.6	5.6	1.4	3	<i>virginica</i>
7.7	3	6.1	2.3	3	<i>virginica</i>
6.3	3.4	5.6	2.4	3	<i>virginica</i>
6.4	3.1	5.5	1.8	3	<i>virginica</i>
6	3	4.8	1.8	3	<i>virginica</i>
6.9	3.1	5.4	2.1	3	<i>virginica</i>
6.7	3.1	5.6	2.4	3	<i>virginica</i>
6.9	3.1	5.1	2.3	3	<i>virginica</i>
5.8	2.7	5.1	1.9	3	<i>virginica</i>
6.8	3.2	5.9	2.3	3	<i>virginica</i>
6.7	3.3	5.7	2.5	3	<i>virginica</i>
6.7	3	5.2	2.3	3	<i>virginica</i>
6.3	2.5	5	1.9	3	<i>virginica</i>
6.5	3	5.2	2	3	<i>virginica</i>
6.2	3.4	5.4	2.3	3	<i>virginica</i>
5.9	3	5.1	1.8	3	<i>virginica</i>

【以下是 iris 資料集，請自行複製存成 iris.txt 檔】

```

5.1 3.5 1.4 0.2 1
4.9 3 1.4 0.2 1
4.7 3.2 1.3 0.2 1
4.6 3.1 1.5 0.2 1
5 3.6 1.4 0.2 1
5.4 3.9 1.7 0.4 1
4.6 3.4 1.4 0.3 1
5 3.4 1.5 0.2 1
4.4 2.9 1.4 0.2 1
4.9 3.1 1.5 0.1 1
5.4 3.7 1.5 0.2 1
4.8 3.4 1.6 0.2 1
4.8 3 1.4 0.1 1

```

4.3	3	1.1	0.1	1
5.8	4	1.2	0.2	1
5.7	4.4	1.5	0.4	1
5.4	3.9	1.3	0.4	1
5.1	3.5	1.4	0.3	1
5.7	3.8	1.7	0.3	1
5.1	3.8	1.5	0.3	1
5.4	3.4	1.7	0.2	1
5.1	3.7	1.5	0.4	1
4.6	3.6	1	0.2	1
5.1	3.3	1.7	0.5	1
4.8	3.4	1.9	0.2	1
5	3	1.6	0.2	1
5	3.4	1.6	0.4	1
5.2	3.5	1.5	0.2	1
5.2	3.4	1.4	0.2	1
4.7	3.2	1.6	0.2	1
4.8	3.1	1.6	0.2	1
5.4	3.4	1.5	0.4	1
5.2	4.1	1.5	0.1	1
5.5	4.2	1.4	0.2	1
4.9	3.1	1.5	0.2	1
5	3.2	1.2	0.2	1
5.5	3.5	1.3	0.2	1
4.9	3.6	1.4	0.1	1
4.4	3	1.3	0.2	1
5.1	3.4	1.5	0.2	1
5	3.5	1.3	0.3	1
4.5	2.3	1.3	0.3	1
4.4	3.2	1.3	0.2	1
5	3.5	1.6	0.6	1
5.1	3.8	1.9	0.4	1
4.8	3	1.4	0.3	1
5.1	3.8	1.6	0.2	1
4.6	3.2	1.4	0.2	1
5.3	3.7	1.5	0.2	1
5	3.3	1.4	0.2	1
7	3.2	4.7	1.4	2

6.4	3.2	4.5	1.5	2
6.9	3.1	4.9	1.5	2
5.5	2.3	4	1.3	2
6.5	2.8	4.6	1.5	2
5.7	2.8	4.5	1.3	2
6.3	3.3	4.7	1.6	2
4.9	2.4	3.3	1	2
6.6	2.9	4.6	1.3	2
5.2	2.7	3.9	1.4	2
5	2	3.5	1	2
5.9	3	4.2	1.5	2
6	2.2	4	1	2
6.1	2.9	4.7	1.4	2
5.6	2.9	3.6	1.3	2
6.7	3.1	4.4	1.4	2
5.6	3	4.5	1.5	2
5.8	2.7	4.1	1	2
6.2	2.2	4.5	1.5	2
5.6	2.5	3.9	1.1	2
5.9	3.2	4.8	1.8	2
6.1	2.8	4	1.3	2
6.3	2.5	4.9	1.5	2
6.1	2.8	4.7	1.2	2
6.4	2.9	4.3	1.3	2
6.6	3	4.4	1.4	2
6.8	2.8	4.8	1.4	2
6.7	3	5	1.7	2
6	2.9	4.5	1.5	2
5.7	2.6	3.5	1	2
5.5	2.4	3.8	1.1	2
5.5	2.4	3.7	1	2
5.8	2.7	3.9	1.2	2
6	2.7	5.1	1.6	2
5.4	3	4.5	1.5	2
6	3.4	4.5	1.6	2
6.7	3.1	4.7	1.5	2
6.3	2.3	4.4	1.3	2
5.6	3	4.1	1.3	2

5.5	2.5	4	1.3	2
5.5	2.6	4.4	1.2	2
6.1	3	4.6	1.4	2
5.8	2.6	4	1.2	2
5	2.3	3.3	1	2
5.6	2.7	4.2	1.3	2
5.7	3	4.2	1.2	2
5.7	2.9	4.2	1.3	2
6.2	2.9	4.3	1.3	2
5.1	2.5	3	1.1	2
5.7	2.8	4.1	1.3	2
6.3	3.3	6	2.5	3
5.8	2.7	5.1	1.9	3
7.1	3	5.9	2.1	3
6.3	2.9	5.6	1.8	3
6.5	3	5.8	2.2	3
7.6	3	6.6	2.1	3
4.9	2.5	4.5	1.7	3
7.3	2.9	6.3	1.8	3
6.7	2.5	5.8	1.8	3
7.2	3.6	6.1	2.5	3
6.5	3.2	5.1	2	3
6.4	2.7	5.3	1.9	3
6.8	3	5.5	2.1	3
5.7	2.5	5	2	3
5.8	2.8	5.1	2.4	3
6.4	3.2	5.3	2.3	3
6.5	3	5.5	1.8	3
7.7	3.8	6.7	2.2	3
7.7	2.6	6.9	2.3	3
6	2.2	5	1.5	3
6.9	3.2	5.7	2.3	3
5.6	2.8	4.9	2	3
7.7	2.8	6.7	2	3
6.3	2.7	4.9	1.8	3
6.7	3.3	5.7	2.1	3
7.2	3.2	6	1.8	3
6.2	2.8	4.8	1.8	3

6.1	3	4.9	1.8	3
6.4	2.8	5.6	2.1	3
7.2	3	5.8	1.6	3
7.4	2.8	6.1	1.9	3
7.9	3.8	6.4	2	3
6.4	2.8	5.6	2.2	3
6.3	2.8	5.1	1.5	3
6.1	2.6	5.6	1.4	3
7.7	3	6.1	2.3	3
6.3	3.4	5.6	2.4	3
6.4	3.1	5.5	1.8	3
6	3	4.8	1.8	3
6.9	3.1	5.4	2.1	3
6.7	3.1	5.6	2.4	3
6.9	3.1	5.1	2.3	3
5.8	2.7	5.1	1.9	3
6.8	3.2	5.9	2.3	3
6.7	3.3	5.7	2.5	3
6.7	3	5.2	2.3	3
6.3	2.5	5	1.9	3
6.5	3	5.2	2	3
6.2	3.4	5.4	2.3	3
5.9	3	5.1	1.8	3