**1. Feature Selection**

3. \*Which sized subset, and which set of attributes gets the best accuracy?\* [1 mark]

|  |  |  |  |
| --- | --- | --- | --- |
| **Subset size** | **Attributes Selected** | **Accuracy** | **Attributes Removed** |
| 5 | All: W, P, Hol, Vac, Health | 85.9649 % | None |
| 4 | W, P, Hol, Health | 89.4737 % | Vac |
| 3 | W, P, Hol | 91.2281 % | Health |
| 2 | W, P | 85.9649 % | Hol |
| 1 | W | 80.7018 % | P |

14 tests were done.

4. \*How many and which attributes are selected? Do they match the results from

Q3?\* [1 mark]

Selected attributes: 1,2,3 : 3

wage-increase-first-year

pension

statutory-holidays

They do match the previous results.

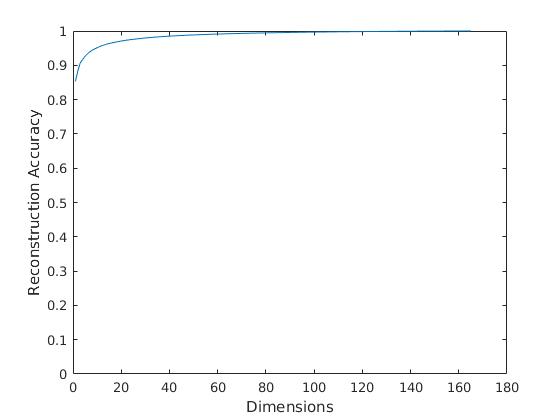
5 \*Which attributes does it pick (and hence which are discarded?)\* [1 mark]

Selected attributes: 2 : 1

sepalwidth

**2. Dimensionality Reduction**

8. \*Use the data cursor on the plot to find out what # of PCs is required to explain 99% of the data variance (achieve 99% reconstruction accuracy). What # is this and does it match the value from Q2.6?\* [1 mark]



We need more than 56 dimensions to achieve 99% reconstruction accuracy.

It matches the value from Q2.6:

Original/Encoding Size: 5407/74 KB. Reconstruction Error: 0.010

9. \*Which number of PCA dimensions gets maximum face recognition accuracy? Is

it better or worse than accuracy classifying the raw images? \* [2 marks]

**27** number of PCA get maximum face recognition accuracy.