Pattern Recognition and Machine Learning

James L. Crowley

ENSIMAG 3 MMIS First Semester 2010/2011

Lessons 1 - 5 January 2011

Exercise:

You are given the weight (kg), height (cm), age (years) and gender for 20 000 people. Weight (kg), Height (cm), Age (years) are given as integers in the following ranges: Weight W [11, 210], Height is from H [121, 220], Age is from A[1, 100]. Gender, (G) is given as the symbol M or F. The data combines information from people from Paris, Chicago, Shanghai and Peru. You are not given the origin of the data.

1. You are asked to construct a 4 Dimensional Table frequencies h(W,H,A,G) for this data. Assuming M=16,000 samples, how many cells are reasonable for h(). Define a "reasonable" quantization for each variable so that your table has this number of cells. Define the transformations for W, H, and A for this quantization.

2. Consider that each cell gives the frequency of occurrence for a specific value (W=w)^(H=h)^(A=a)^(G=g). For notational convenience, let us write this as W^H^A^G

Give formulas for P(W^H^A), P(W^H), P(W).

3. Show that P(W^H|A) = P(A|W^H) P(W^H)/P(A)

4. Show that P(W) = P(H) P(W | H) / P(H |W)

5. Explain how to use EM to discover 4 distinct Gaussian functions to represent the data.