Four Dimensional Barycentric Plots in 3D

Geoffrey Matthews
Western Washington University
Bellingham, Washington, USA

February 17, 2006

An n-dimensional vector that describes a probability distribution over n possible outcomes, such as $\langle p_1, p_2, \ldots, p_n \rangle$, is overdetermined because of the requirement $\sum_i p_i = 1$. Hence points of a two-class distribution can be plotted on the line $[0, 1]$, with the endpoints representing the distributions $(0, 1)$ and $(1, 0)$, and an intermediate point, such as 0.3, representing $(0.3, 0.7)$. Points from a three-class distribution can be plotted over a triangle, with the corners of the triangle representing the distributions $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. Such plots (an example is seen in Figure 1, from the **triplot** help file in the **klaR** package [3]) are quite common in numerous disciplines.

Points over a four-class distribution can be plotted as points within a three-dimensional tetrahedron, with the four corners of the tetrahedron representing the distributions $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, *etc.* Such visualizations are useful in clustering and classification problems where points are given probabilistic membership in classes, and a 2D plotting method is given in the **klaR** package [3]. An example figure from the help page of the **klaR** package is provided as Figure 2.

I have developed a package, **quadplot3d**, using the **rgl** package [1] to create genuine 3D plots of four dimensional points in the probability distribution subspace. Figure 3 shows how the plots from Figure 2 look when replotted in 3D. The user can, of course, interact with the 3D figures, rotating, scaling, and changing the background, as with any other **rgl** plot.

In addition to plotting points, some of the features of the **misc3d** package [2] have also been ported to **quadplot3d**. Isosurfaces of a four dimensional function, for example, can be produced by the **quadcontour3d** routine, mimicking the **contour3d** function of the **misc3d** package. In Figure 4 I have plotted isosurfaces (at values 0.7, 1.5, and 1.8) of the *entropy* function, $-\sum_i p_i \log p_i$. The entropy function is widely used in data mining applications. An intuitive intuitive understanding of entropy, the placement of its troughs and ridges, for example, can be gained from the figure.

Visualizing points and functions in four dimensions has proved an important tool in my understanding of data as well as of the clustering, classification, and data mining algorithms that deal with probabilistic class membership. I believe this tool will provide a valuable adjunct to packages such as **klaR**, that attempt to bring together statistical and machine learning approaches. **R** provides a good platform for this kind of cross-disciplinary research.

References

[1] Daniel Adler. *rgl: 3D visualization device system (OpenGL)*, 2004. R package version 0.64-13.

[2] Dai Feng and Luke Tierney. *misc3d: Miscellaneous 3D Plots*. R package version 0.3-1.

[3] Claus Weihs, Uwe Ligges, Karsten Luebke, and Nils Raabe. klar analyzing german business cycles. In D. Baier, editor, *Data Analysis and Decision Support*, pages 335–343, Berlin, 2005. Springer-Verlag. (in print).

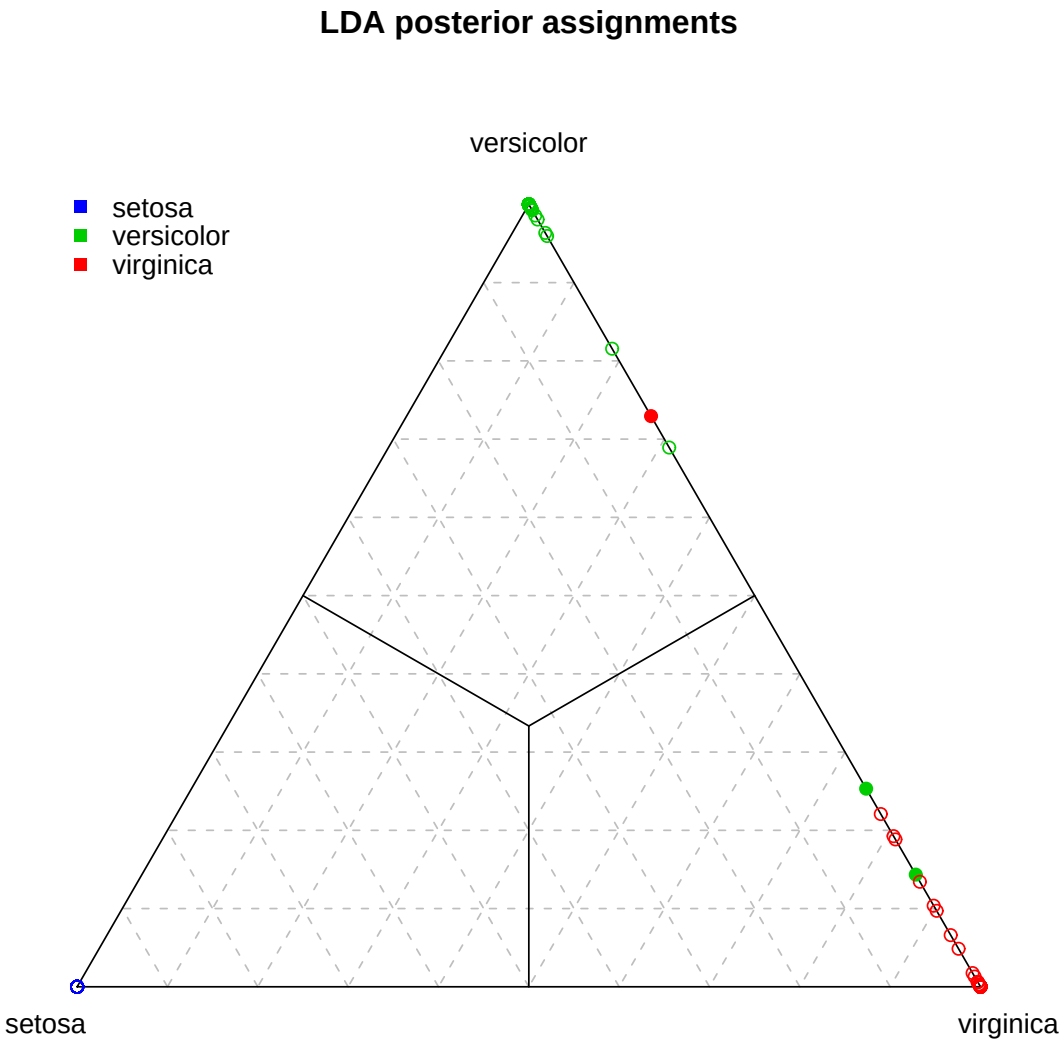

Figure 1: Barycentric plot of three dimensional points as provided by the **triplot** function of **klaR** package. Linear discriminants analysis (LDA) has provided a probability for each point's membership in the three classes of iris. Probability of membership in each of the three classes is given by distance from the "corner" positions, which represent unambiguous membership in the named class.

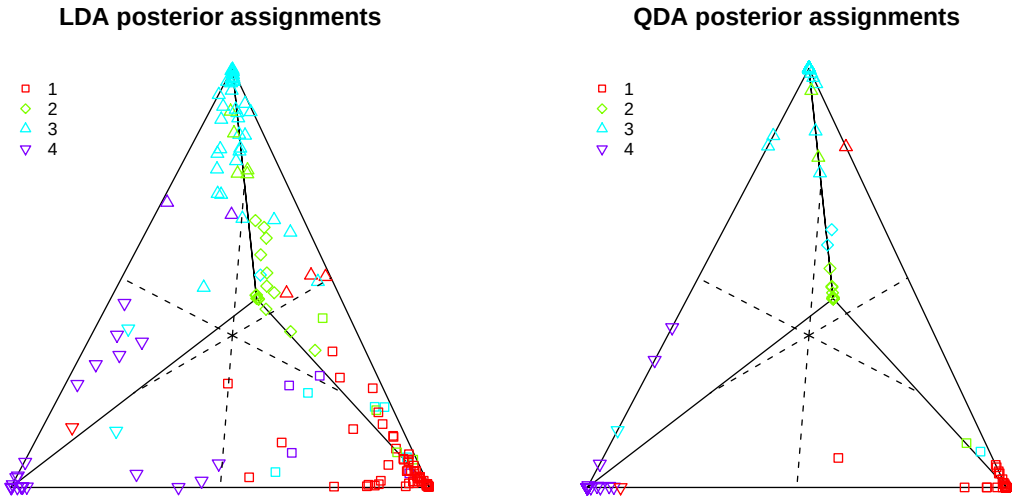

Figure 2: Two dimensional barycentric plots of four dimensional points as provided by the **quadplot** function of the **klaR** package. Here linear discriminants (LDA) and quadratic discriminants (QDA) have provided probabilistic membership for each of four classes. The corners in this case are the corners of a three dimensional regular tetrahedron, which has been projected down into two dimensions for the plot. The difficulties of visualizing four dimensions in a 2D projection are apparent.

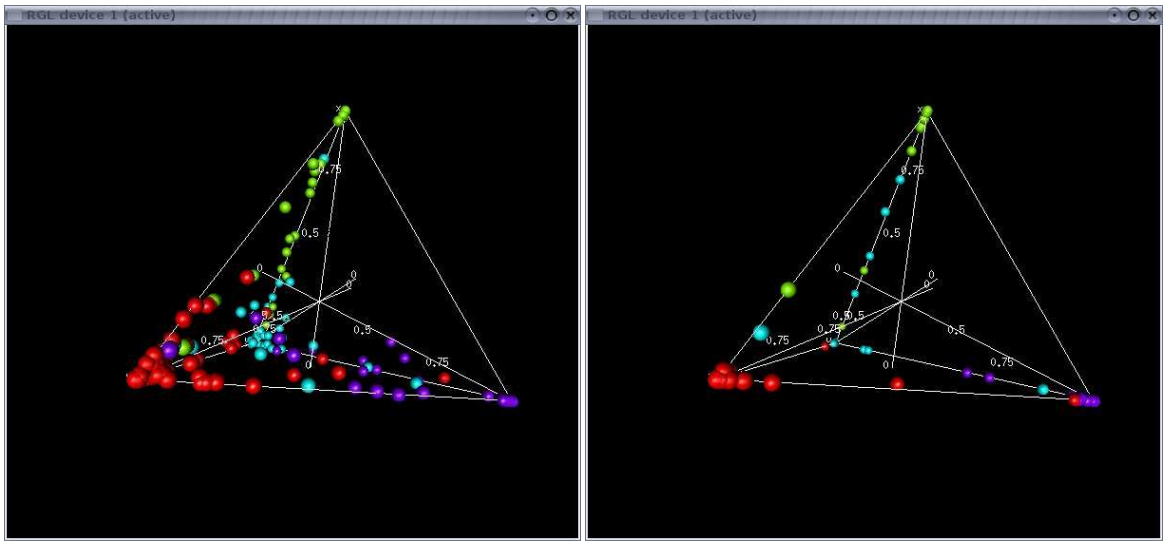

Figure 3: LDA and QDA posterior assignments in four dimensions (as in Figure 2) plotted in 3D using **quadplot3d**. 3D graphics provided in **R** by the **rgl** package are used to provide an interactive 3D view of the tetrahedron describing class membership. 3D hints such as size and perspective are apparent, and make the true relationships among the data points easier to see. Compare this figure with Figure 2.

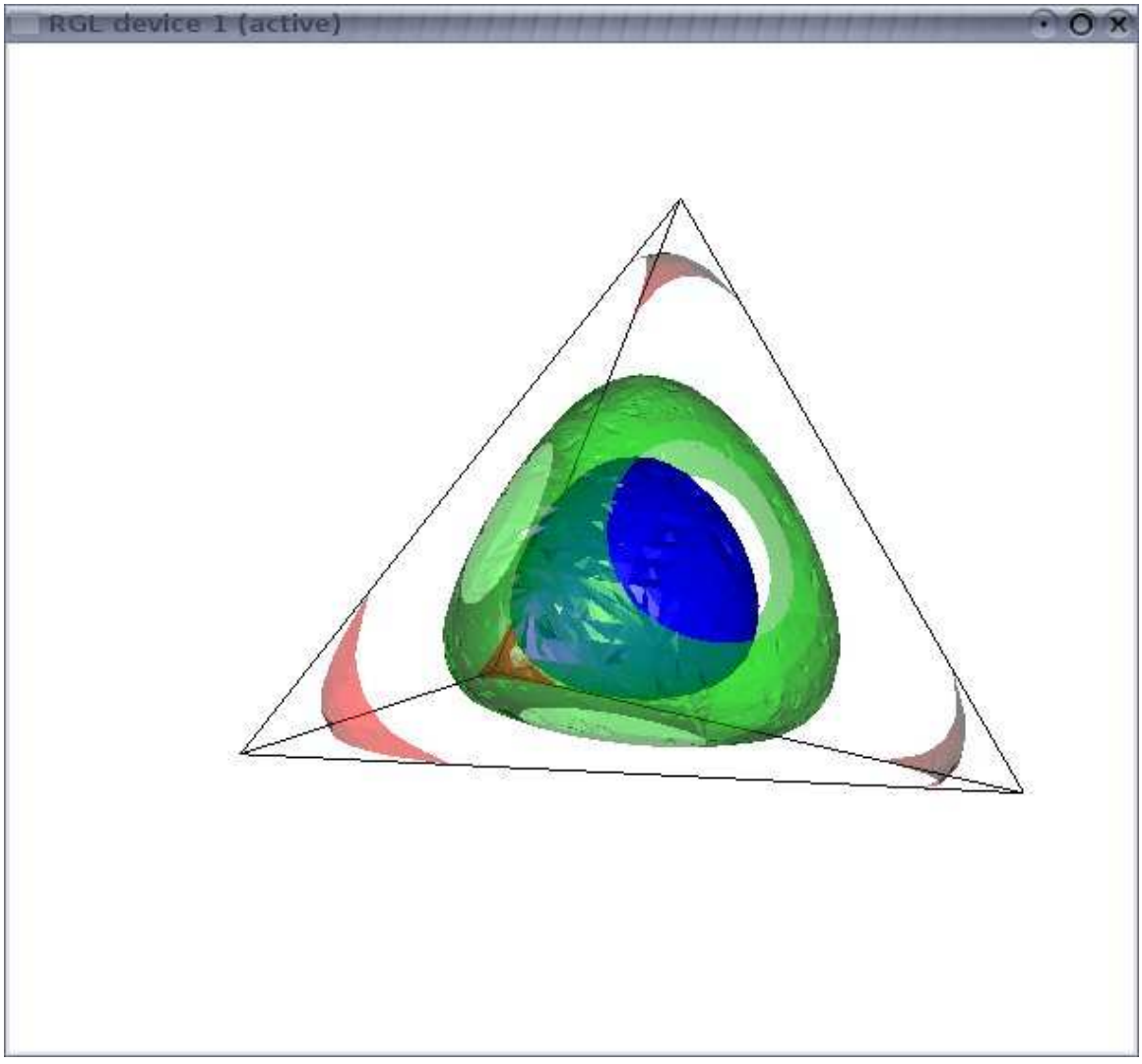

Figure 4: Isosurfaces of the entropy function in four dimensions plotted in 3D using **quadcontour3d**. A simple marching tetrahedra algorithm extracted surfaces of approximately equal value throughout the tetrahedra. Values of entropy used in the figure are 0.7, 1.5, and 1.8.