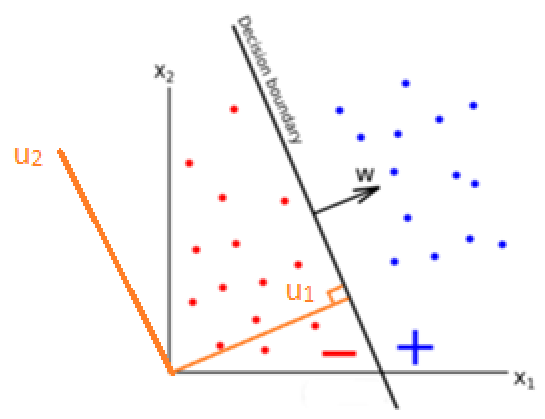
**Linear Discriminant Analysis**

Continuing the classification theme. Say we have two distinct categories in some generalized feature vector space. And we’d like to find some boundary/generalized surface in this space which separates them. To facilitate this, we try to find a new set of orthogonal axes u1,2,3,…,n along which most, second most, third most, etc., of the variation between the categories occurs. In some cases, if the data are separable by some hyperplane, then I imagine u1 would be perpendicular to this hyperplane, as on the left. But if the data are not separable by a plane, then I think u1 would be perpendicular to the plane that does the best job. In neither case, does this formalism yet tell us where that plane is along axis u1?

 Chart, scatter chart

Description automatically generated

(domain boundary and u2 should’ve been more obviously parallel) Note we *could* draw an elaborate surface that exactly separates the two classes, but I think this would be like fitting noise instead of the underlying physics. Seems like we could use Logistic Regression (univariable or multivariable) to do this as well, with some p-cutoff like p = 0.5 or whatever. I’ve read that Logistic Regression often gives better results, but in specific circumstances, LDA is better. So we’ll proceed with LDA.

**2 Categories**

Like with PCA, we are going to conjure up a new orthogonal coordinate system. And we want to find an axis along which the following is maximized.



where:



So we can write:



Minimizing u w/r to **u**, subject to normalization constraint should give us an eigenvalue problem, like with PCA. So diagonalizing the resulting matrix will give us the new axes = eigenvectors, and variations along those axes = eigenvalues. Largest eigenvalue and eigenvector corresponds to u1, next largest to u2, etc.

**3 or more Categories**

If we have three categories that we’re trying to separate, then we’ll minimize,



where is the overall average, projected onto . This doesn’t seem to reduce to the 2 Category formula.