The Geometry of Multicollinearity

David M. Baker, PG
Senior Geological Advisor / Lead Data Scientist
Chesapeake Energy Corp.
October 2016

Abstract

To better understand the effects of multicollinearity on predictive models, the geometry of multicollinearity is developed and illustrated by example. The relationship between the Pearson correlation coefficient and the correlated bilinear geometric forms in three-dimensional space is shown, and the breakdown of the bilinear forms, near positions of linear dependence, demonstrates the deleterious effect of multicollinearity on separability in predictive models.

Discussion

The problem of and the intuition around the multicollinearity between the attributes of a predictive model can be further understood and visualized geometrically as the breakdown of the redundancy in the bilinear forms on the vector space of the model attributes when pairs of feature attribute vectors near the degenerate or singular cases. This condition occurs when any two of these vectors approaching parallel degenerate into the line formed by their vector sum or when approaching the singularity at the anti-parallel position where their vector sum vanishes. When not in these positions, the span of the feature attribute vectors define a (hyper)plane [1]. When in these positions, a degree of freedom is lost leading to an infinite number of planes passing through the parallel or anti-parallel line formed, and no separability in the classes defined by the feature attributes is possible.

The summation vector of the normalized feature attribute vectors not only lies in the plane spanned by, but also bisects the feature attribute vectors in the plane. It is the plane containing the bisector that is also perpendicular to the feature attribute vector plane that maximally separates the classes of the feature attribute vectors, see **Figure 1**. In three-dimensional space, this perpendicular plane is also defined by the span of the cross product vector of the feature attribute vectors and the bisector [1].

The closer to orthogonal or, equivalently, the greater the angle between the pair of feature attribute vectors the greater is the discriminatory power of the attributes to the model [2]. Important to feature evaluation is the correlation between the features of the model. The Pearson product-moment correlation coefficient is classically defined as a function of the raw scores and the means of the variable values of two features [3]. Geometrically, and when the feature attribute vectors are centered, the cosine of the angle between the vectors is also equal to the correlation coefficient. This definition is also the dot or scalar product between the two vectors; that is, the dot product of two centered feature attribute vectors equals the Pearson product-moment correlation coefficient between the attributes [3].

The paragraphs above detail three bilinear forms, see **Figure 1**. Given two vectors in 3D space, **u** and **v**, and the scalar values, ρ and σ, the linear combination of the scaled vectors is $\rho\mathbf{u} - \sigma\mathbf{v} = \mathbf{r}$, such that **u**, **v** and **r** will, by construction, all lie in a common plane Π, that common plane defined by a pair of any two of these three vectors, as long as the vector pair is neither parallel or anti-parallel to each other, the degenerate and singular cases we consider here. The dot or scalar product of two vectors equals zero, $\mathbf{u} \bullet \mathbf{w} = 0$, only when these two vectors are orthogonal to each other. Finally, the cross product of the pair of vectors **u** and **v** is the vector **w**, perpendicular to the plane spanned by the vectors, Π, and is also orthogonal to each vector of the pair $\mathbf{u} \times \mathbf{v} = \mathbf{w}$. The three previous forms are so related and so well correlated that [1]:

$$(\rho\mathbf{u} - \sigma\mathbf{v})\bullet(\mathbf{u} \times \mathbf{v}) = (\mathbf{r} \bullet \mathbf{w}) = 0.$$

Assuming arithmetic calculated using infinite precision, this correlation says that for all values, ρ, **u,** σ, **v**, and for any and all pairs of vectors in the plane spanned by a pair, the cross product of those vectors is a vector orthogonal to the pair and perpendicular to the plane defined by them [4]. But, these bilinear forms may be so affected by error in the data, and round-off error and cancellation in floating point arithmetic, that near singularities and when approaching degenerate positions, the expected geometric redundancy in the correlated operations is compromised to such an extent that the results are no longer valid for the given accuracy of the data [4].

To visualize the effect of the compromised geometric form and its relation to the multicollinearity, first recall that in linear models, classes are separated by hyperplanes [5]. Three-dimensional space will be used for illustration. From the description above, the two spanning vectors <u>not</u> in degenerate or singular position, here called normal position, define a plane. The cross product vector, that one normal to the plane, is also classically used to define the orientation of the plane. Assume only one of the spanning vectors has any error. The error in the direction of the other spanning vector may be visualized as the vector falling anywhere at random inside a cone of maximum error while rotating around the true direction of the vector. In normal position, visualize, as the error vector moves around inside the cone of error, how the orientation of the plane and the direction of the normal to the plane changes. If the cone of error is relatively small, though the orientation and the direction of the normal vector change, the general direction of the normal is still the same.

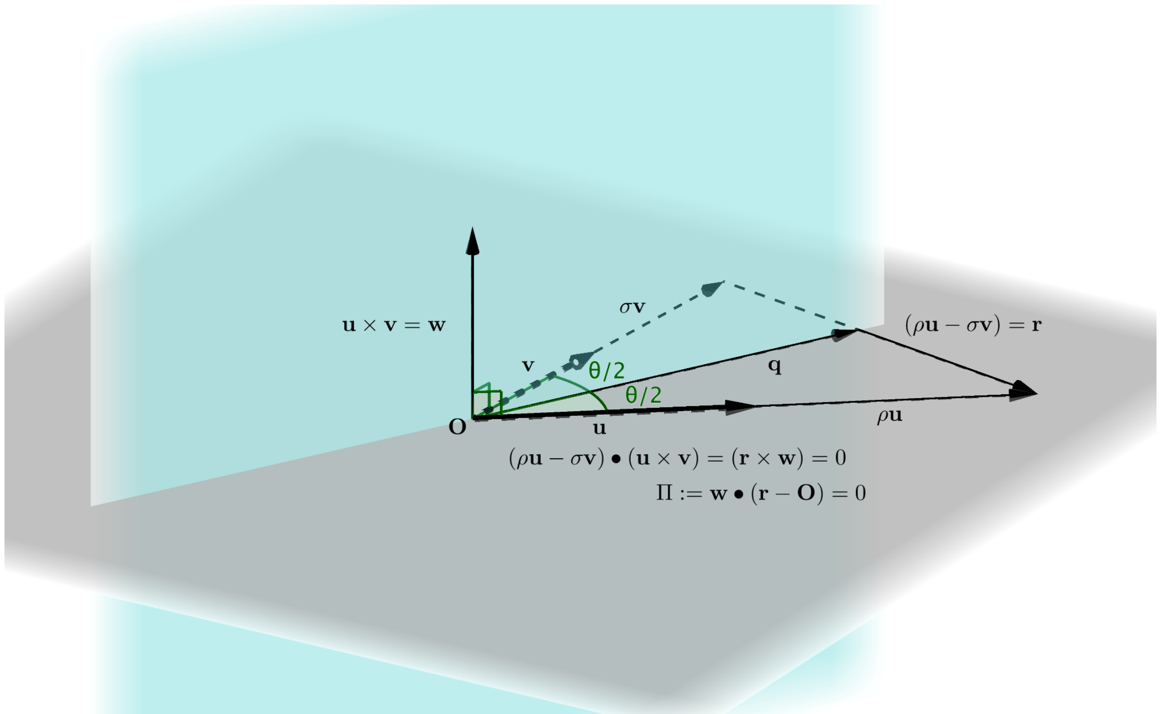

Figure 1: Bilinear forms. The plane Π spanned by the vectors **u** and **v** is defined by the vector **w,** the cross product of **u** and **v**, and position **O** on the plane. After the linear combination on **u** and **v,** ρ**u** and σ**v**, the difference vector **r** dotted with **w** is equal to zero since **r** lies in the plane and **w** is orthogonal to the plane, demonstrating the geometric redundancy of the forms. The plane spanned by the angle bisector **q** and **w** is the maximally separating plane.

Now consider the converse position, where the position of the spanning vectors are near (anti)parallel. In this position, it is possible that the small changes in the error vector, as it rotates around its true direction, can cause very large and rapid changes in the direction of the normal vector to such an extent the redundancy of the bilinear forms no longer holds. The stability of the bilinear form asymptotically vanishes as the spanning vectors approach collinearity.

Illustrative Example

To make it concrete, imagine a dancer, dancing in place, her legs are the vectors of the plane the normal vector pointing forward, see **Figure 2**. With slight abuse of the geometry, draw two small circles on the dance floor shoulder width apart that are the bases of cones of error, and the dancer's feet at all times must stay within the circles. In this situation, no matter where the dancer's feet are, she will be always looking in the same direction. The dancer's legs are in normal position, and the spectators will always be in front and backstage in back separating the spectator class from the backstage staff class.

Next consider a ballerina dancing on her toes, her ankles nearly touching. She too must keep her toes within the bounds of her error circles, these circles almost fully overlapping. While tapping on her toes, she can easily and quickly move around in a circle or change direction with just a few small steps, all the while keeping her toes in their respective error circles. Because of the violation of the redundancy of the bilinear forms, the ballerina's direction no longer can be used to separate upstage from backstage.

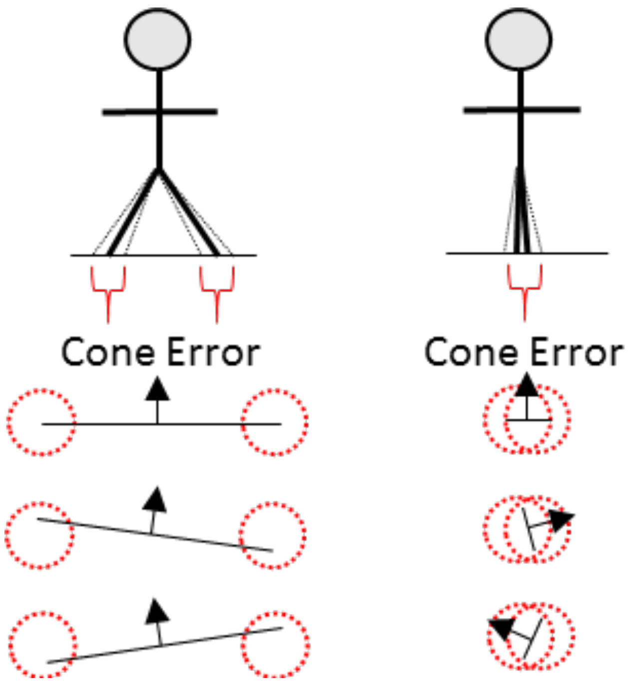

Figure 2: Left, a dancer dancing in place. Legs are in normal positon so small changes in foot position cause only a small changes in direction. In general she will still be facing approximately the same direction. Right, the ballerina on tip toes. Here, where her legs are nearly parallel, a small change in toe position can cause a very large change in direction. A few taps of the toe and she may be facing a completely different direction.

Conclusion

If there were never any error in the direction of the spanning vectors, then only when the vectors become linearly dependent at the extremes, the degenerate and singular cases, would there be infinitely many planes and no separation in classes possible. But all predictive models have error, and the multicollinearity problem occurs when two or more feature attribute vectors point nearly in the same or opposite directions; they are well correlated. As such, when feature attribute vectors are highly correlated, collinear, the orientation of the class-separating hyperplane becomes unstable because of the violation of the strong correlation of the bilinear geometric forms. Thus, the effects of multicollinearity on a predictive model can be seen geometrically in the quality of the bilinear forms as they degrade asymptotically when the degenerate and singular positions are approached.

References

[1] H. S. M. Coxeter, Introduction to geometry, New York: Wiley, 1989.

[2] Wikipedia, "Multicollinearity," [Online]. Available: https://en.wikipedia.org/wiki/Multicollinearity. [Accessed 8 aug 2016].

[3] Z. Gniazdowski, "Geometric interpretation of a correlation," *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki,* vol. 9, no. 7, pp. 27 - 35, 2013.

[4] W. Kahan, "Cross-Products and Rotations in 2-and 3-Dimensional Euclidean Spaces," 2 mar 2008. [Online]. Available: https://www.cs.berkeley.edu/~wkahan/MathH110/Cross.pdf. [Accessed 31 dec 2015].

[5] Wikipedia, "Decision boundary," [Online]. Available: https://en.wikipedia.org/wiki/Decision_boundary. [Accessed 8 aug 2016].