

ALGEBRAIC MODEL SELECTION AND IDENTIFICATION USING GROEBNER BASES

BRANDILYN STIGLER

ABSTRACT. Developing predictive models from large-scale experimental data is an important problem in biological data science. In certain settings, it may be advantageous to discretize the data to improve the signal-to-noise ratio or reduce data size, for example. While there are many classes of functions that can fit data from an underlying network, all data discretized over a finite field are fit by polynomials. A useful consequence is that polynomial models of such data can be written in terms of a standard monomial basis, by way of Groebner bases of zero-dimensional ideals, where each choice of standard monomials provides a different prediction regarding network structure.

In this work, we use affine transformations to partition data sets into equivalence classes with the same sets of standard monomials. This partition reveals that data sets with unique bases are a so-called linear shift of a staircase. Implications of this work are guidelines for designing experiments which maximize information content and for determining data sets which yield unambiguous predictions.