

Random projections are a useful technique for reducing the computational complexity of routine data analysis computations. We focus on their application to ordinary least squares regression, a benchmark predictive technique that can be computationally expensive on massive datasets. Sketched regression algorithms use random projections to compress the original dataset to a manageable size. Sketched regression algorithms have a number of highly desirable algorithmic properties, particularly in terms of worst case performance. We make a new connection between sketched regression and statistical linear modelling. The linear modelling framework reveals new statistical properties of sketched regression algorithms. The statistical view complements the existing theoretical framework for sketched regression and leads to analytical results about the accuracy of the procedure.