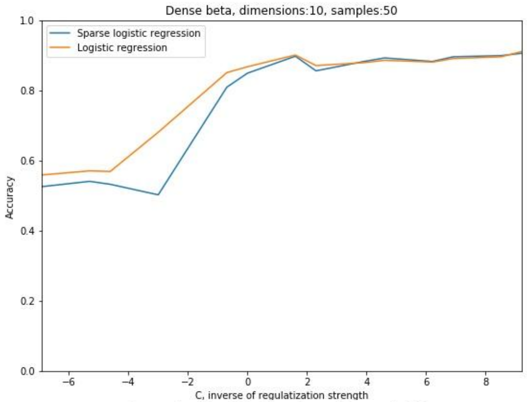
Sparse vs dense classification

Generating data and models used

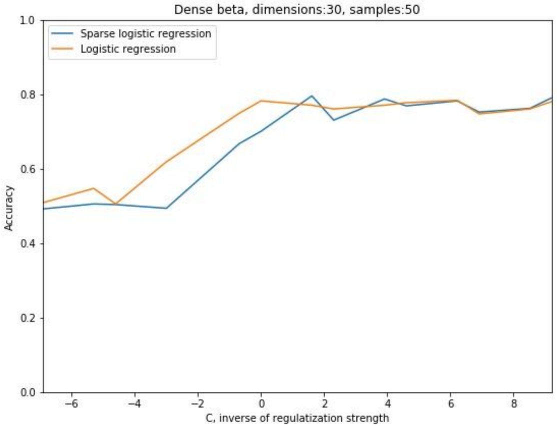
- Data generated according to $y=X\boldsymbol{\beta}+\varepsilon$ where $X \in R^{(n \times p)}$, $y \in R^n$, $\boldsymbol{\beta} \in R^p$, $\varepsilon \in R^n$ where n is the number of samples and p is the number of features/dimensions.
- $X \sim N(0,I\boldsymbol{\sigma}^2)$ where I is the identity matrix and sigma squared is 1/sqrt(dimensions).
- $\boldsymbol{\beta} \sim N(0,I)$ where I is the identity matrix. For testing sparse data, a number of entries in $\boldsymbol{\beta}$ was set to 0.
- $\varepsilon \sim N(0,I\boldsymbol{\sigma}^2)$ is the noise and $\boldsymbol{\sigma}^2$ was set to a value such that the sound to noise ratio took values between 1 and 60.
- The labels to y was assigned through coupling a sample from X to the sign of the output y to get a binary classification
- The models compared are
  1. Logistic Regression and Sparse Logistic Regression
  2. Naive Bayes/Nearest Centroids and Shrunken Centroids
- All plots is the average over 50 trials where new data is generated every trial

Logistic regression and Sparse logistic regression

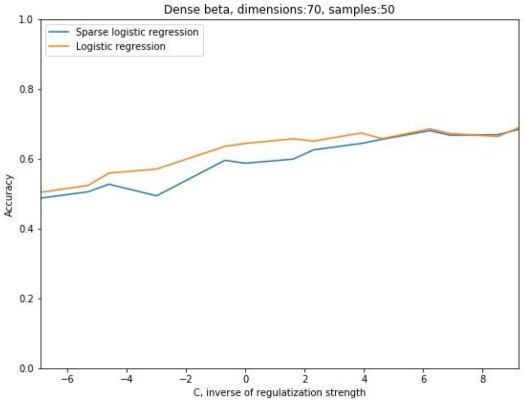
Lots of parameters to vary, here we compare dense beta vs sparse beta for varying dimensions and C, SNR=20



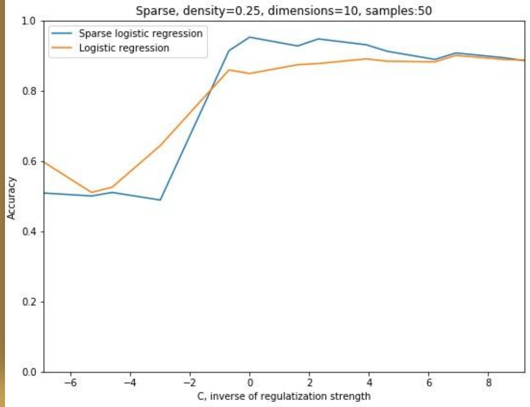
Dense beta, dimensions:10, samples:50



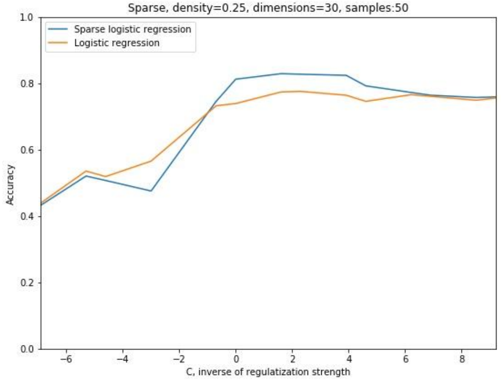
Dense beta, dimensions:30, samples:50



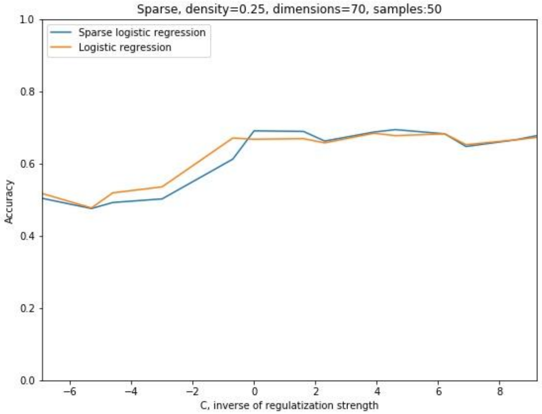
Dense beta, dimensions:70, samples:50



Sparse, density=0.25, dimensions=10, samples:50



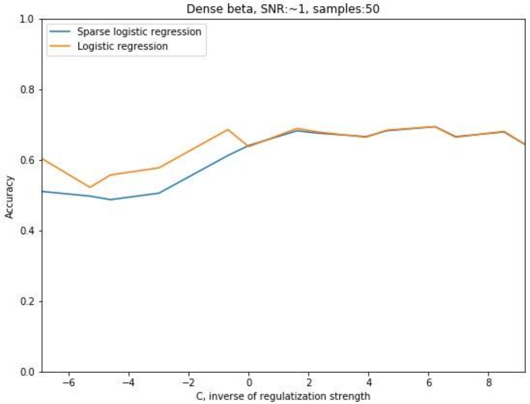
Sparse, density=0.25, dimensions=30, samples:50

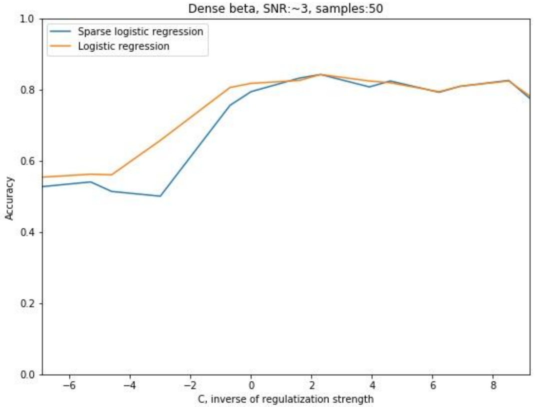


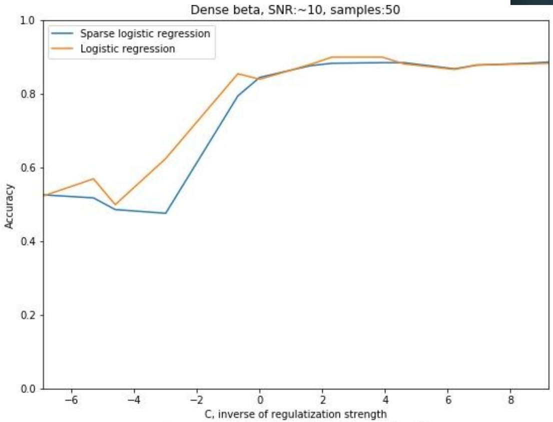
Sparse, density=0.25, dimensions=70, samples:50

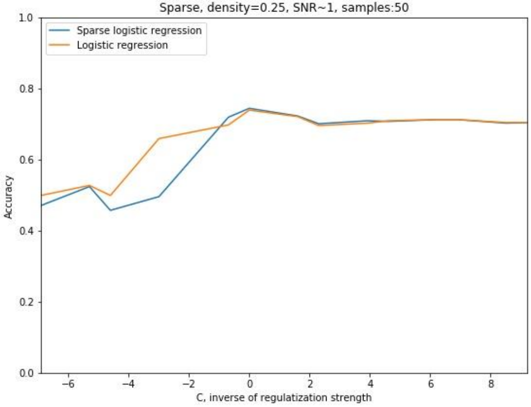
Comparing low noise and high noise, dim=10

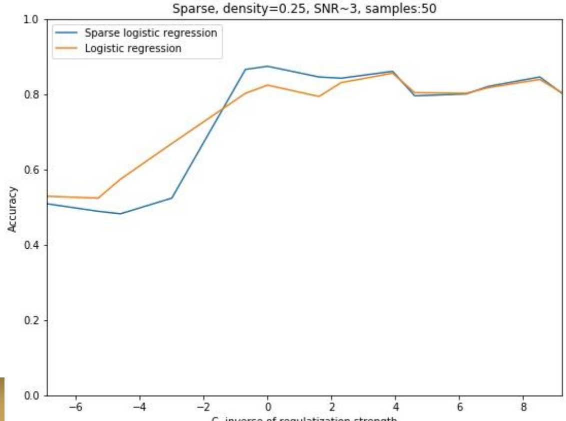
- No perceivable difference between low or high noise levels for different densities

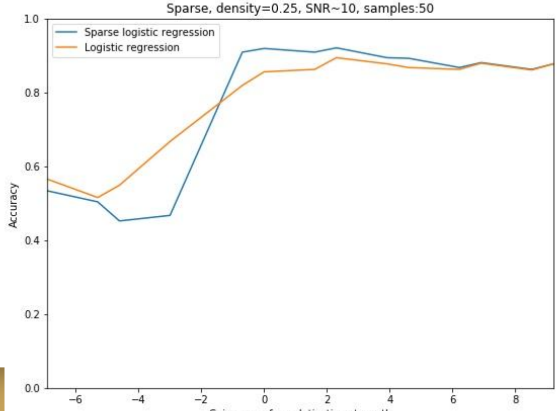










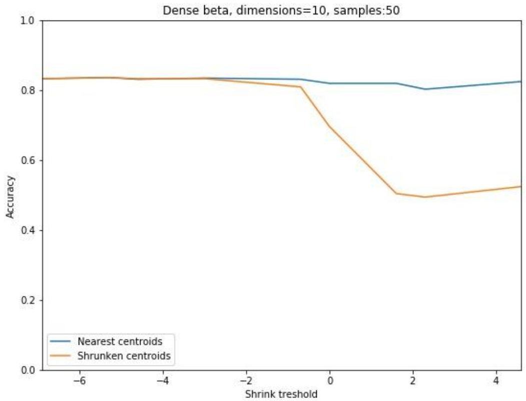


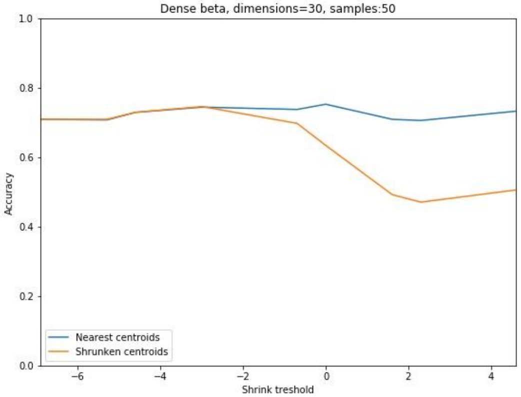
Results and conclusions

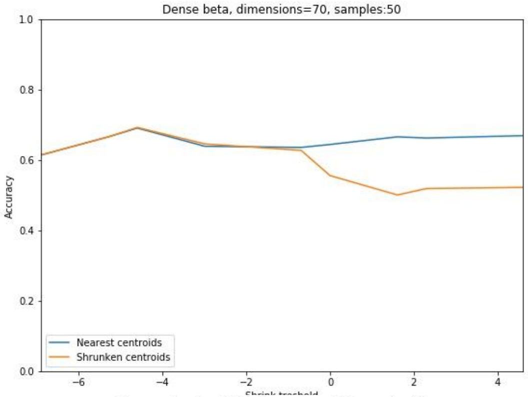
- The dimensionality didn't have an effect on which method was best according to these results, but we saw that the sparse method (lasso) had a higher accuracy score when $\boldsymbol{\beta}$ was sparse than when it was dense. This makes sense since lasso assigns 0s to some of the weights in $\boldsymbol{\beta}$ when fitting them, and since many entries in $\boldsymbol{\beta}$ actually were 0 when the data was generated this is a viable simplification.
- For sparse data sets the sparse logistic regression had a higher score than the Logistic regression for varying dimensionalities and noise levels (and for the right value of lambda)
- The noise level didn't seem to have as big of an impact as sparsity on which method scored higher, hard to say if the difference was due to difference in noise level or density
- Hard to say which results come from varying one parameter and which results comes from parameters working together so to say

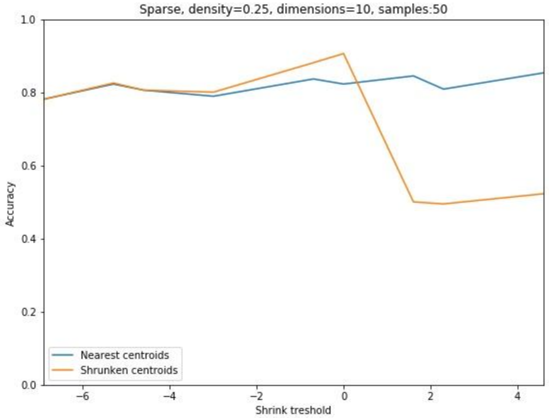
Nearest centroids vs Shrunken centroids

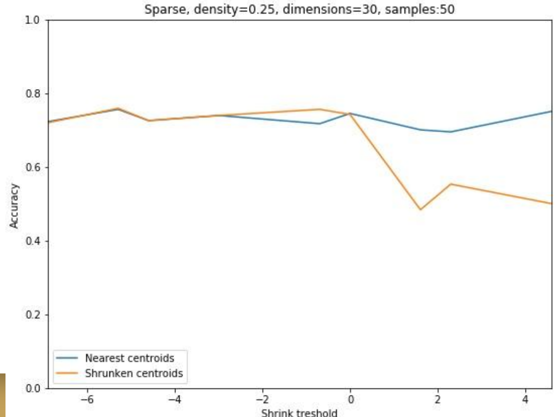
Lots of parameters, here we compare dense beta vs sparse beta for varying dimensions and shrink treshold, SNR=30

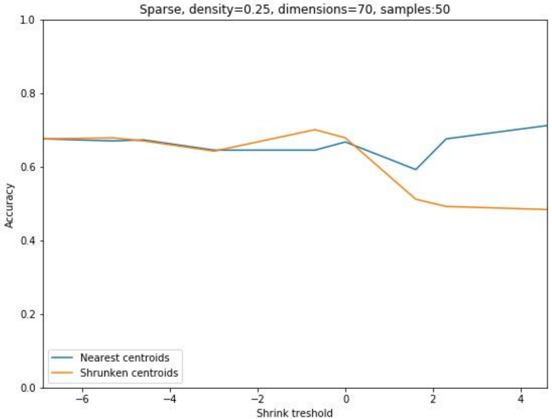

Dense beta, dimensions=10, samples:50


Dense beta, dimensions=30, samples:50


Dense beta, dimensions=70, samples:50

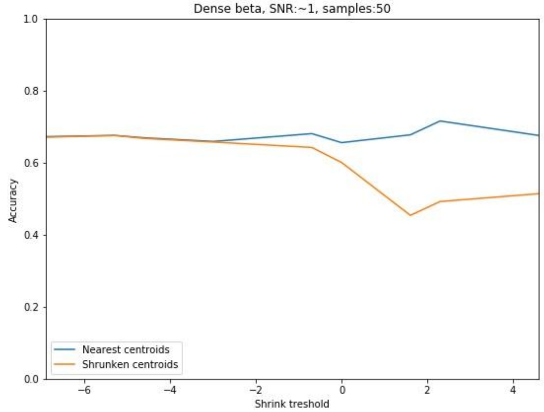

Sparse, density=0.25, dimensions=10, samples:50

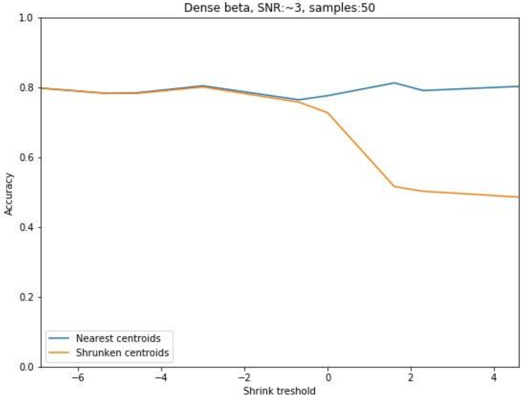

Sparse, density=0.25, dimensions=30, samples:50

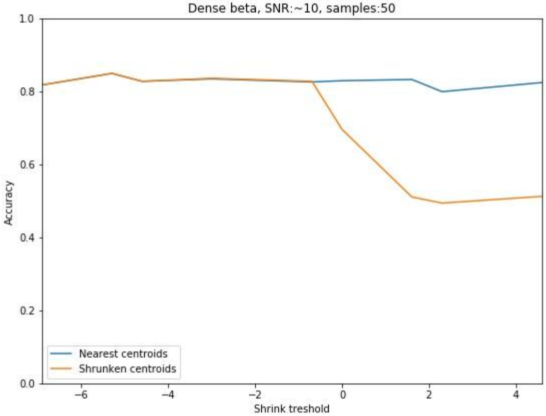

Sparse, density=0.25, dimensions=70, samples:50

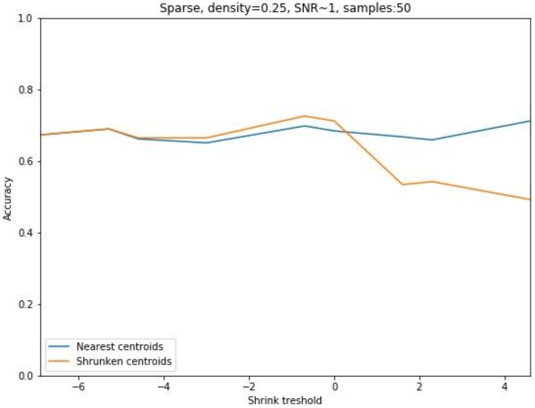
Comparing low noise and high noise, dim=10

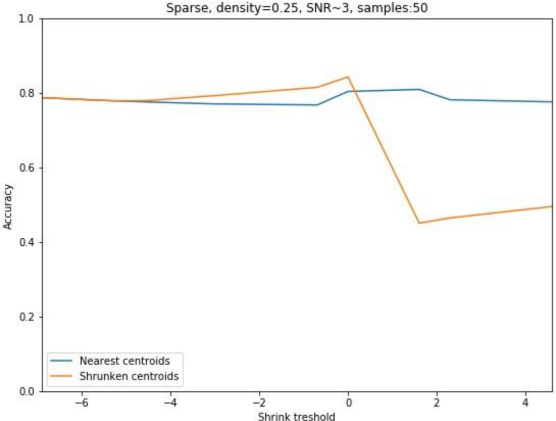
Not much impact from varying the noise, other than decreasing accuracy

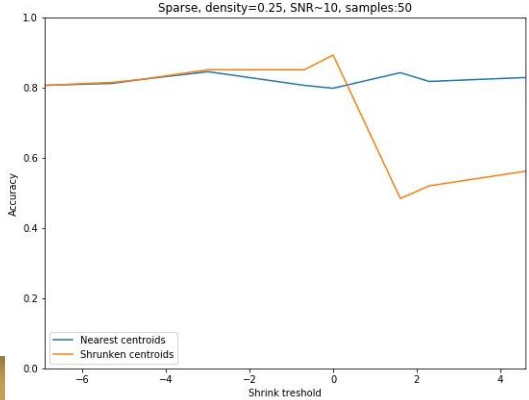












Results and conclusions

- The dimensionality didn’t have an effect on which method was best according to these results
- For sparse data sets the shrunken centroids method was better than the nearest centroids method for varying dimensionalities and noise levels. This makes sense since the shrunken centroids method discard some of the dimensions of our data, i.e setting weights in $\boldsymbol{\beta}$ to 0, which is a close approximation of reality since the data was generated with a $\boldsymbol{\beta}$ with some weights set to zero.
- The noise level didn’t seem to have as big of and impact as sparsity on which method was better, hard to say if the difference was due to difference in noise level or density
- Hard to say which results come from varying one parameter and which results comes from parameters working together so to say