Predicting contrast performance for the Gemini Planet Imager

Victoria Borish, Jean-Baptiste Ruffio, Katherine Sytwu

The Gemini Planet imager is a high contrast imager meant to discover and characterize new exoplanets.

Our goal is to help astronomers decide whether to observe under certain environmental conditions by predicting the contrast value.

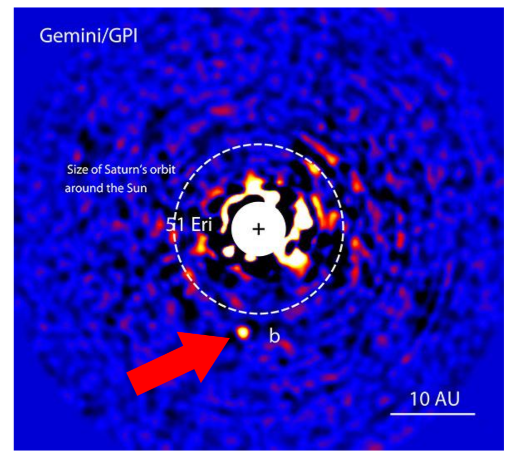


Figure 1: Newly discovered exoplanet 51 Eridani b.

Features and Targets

The **contrast** is the relative brightness of the faintest viewable exoplanet. Therefore, the smaller the better.

We've selected a subset of environmental features measured by the observatory. Our raw data set has 1423 training examples and 7 features.

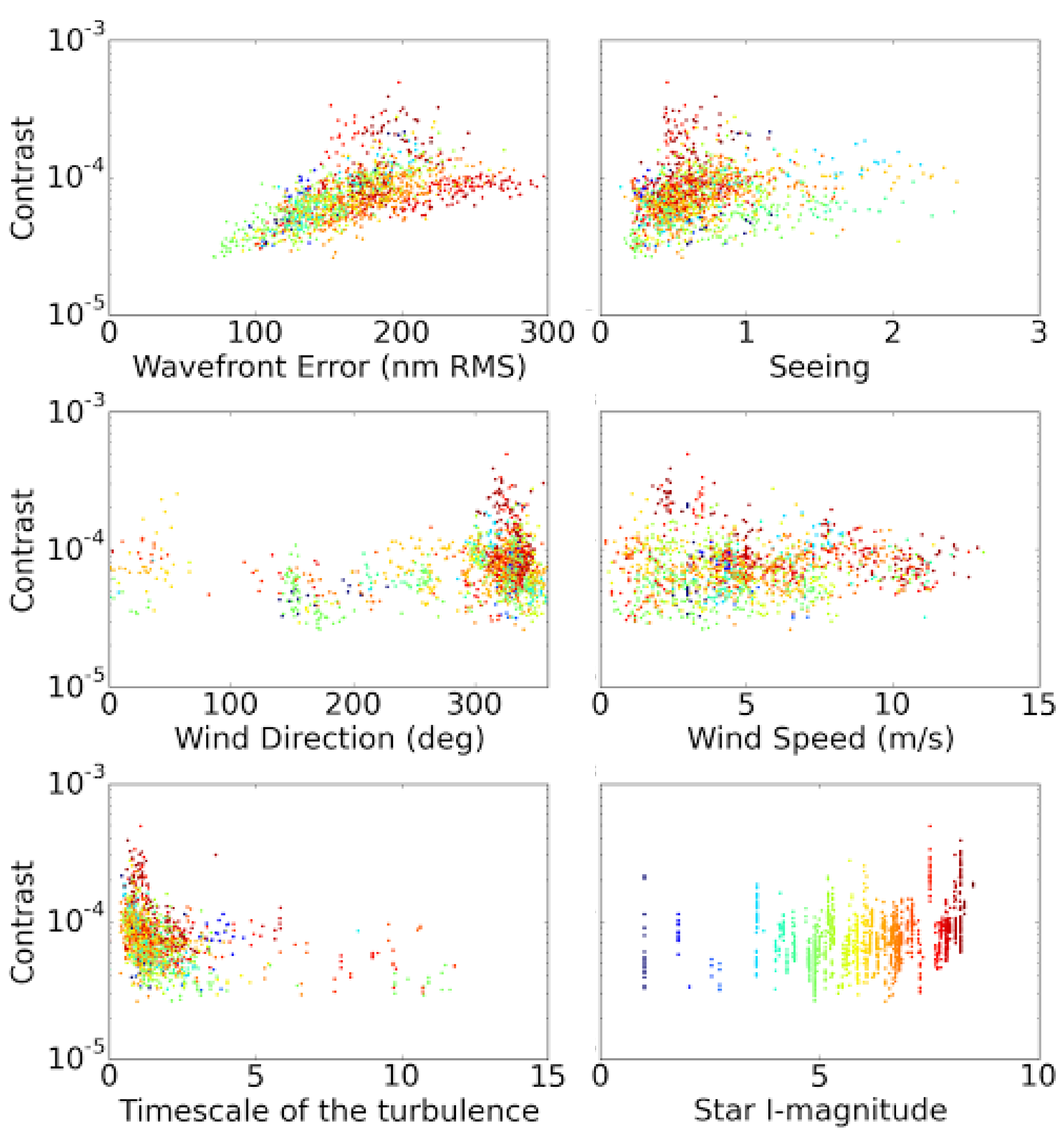
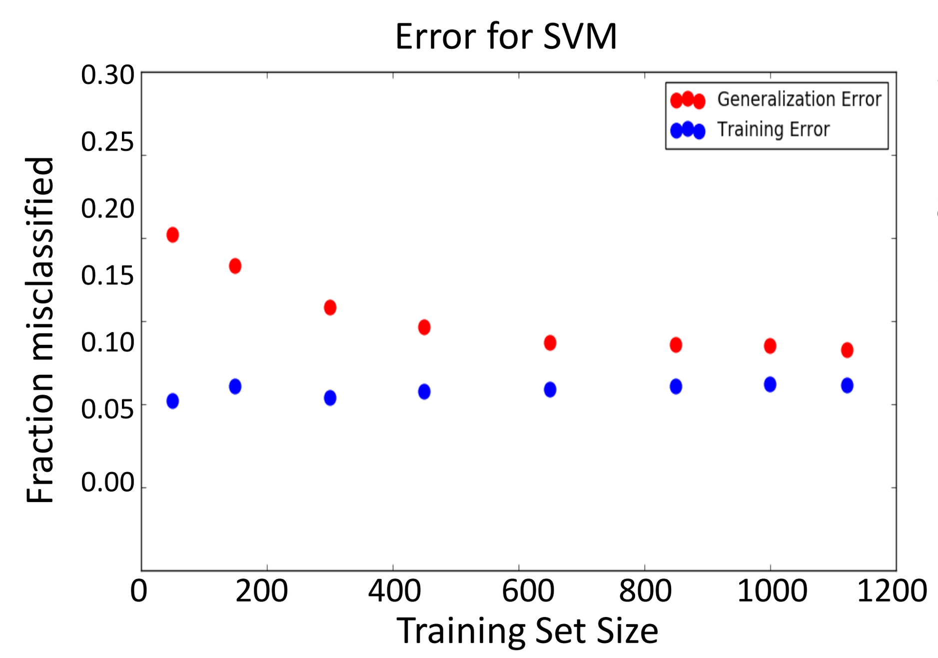


Figure 2. Scatter plots of some features against the contrast for the raw data set.

Should we observe tonight (classification)?

Support vector machines can handle complex problems and can easily switch between kernels.

$$\min_{\gamma,w,b} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$$
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i,$$
$$\xi_i \geq 0, \quad i = 1, \ldots, m.$$



With the raw data, the generalization error levels off at around 12.6%.

Problem: data is too noisy

Figure 3. Generalization and training error for SVM for various training set sizes averaged 10 times.

Solution: Process the data, adjust penalty parameter C

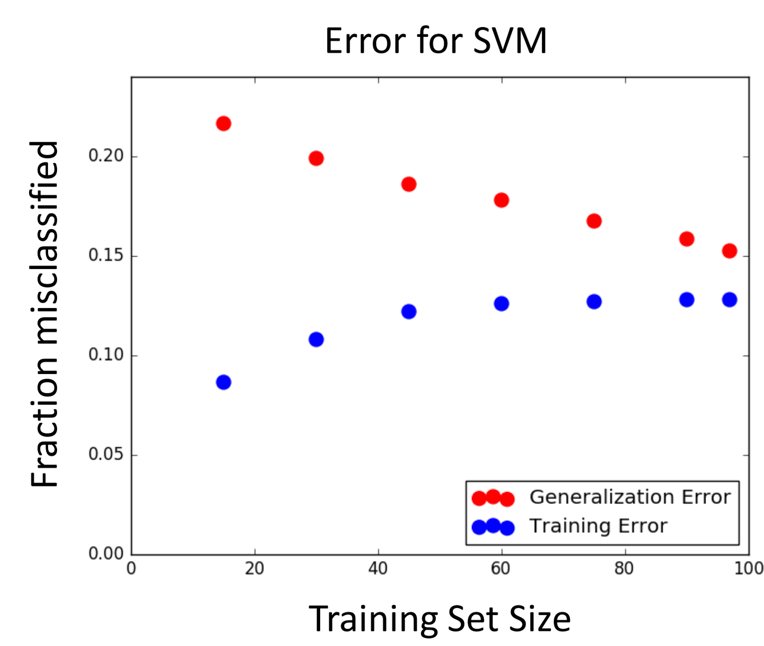


Figure 4. Generalization and training error for SVM with processed data for various training set sizes averaged 1000 times.

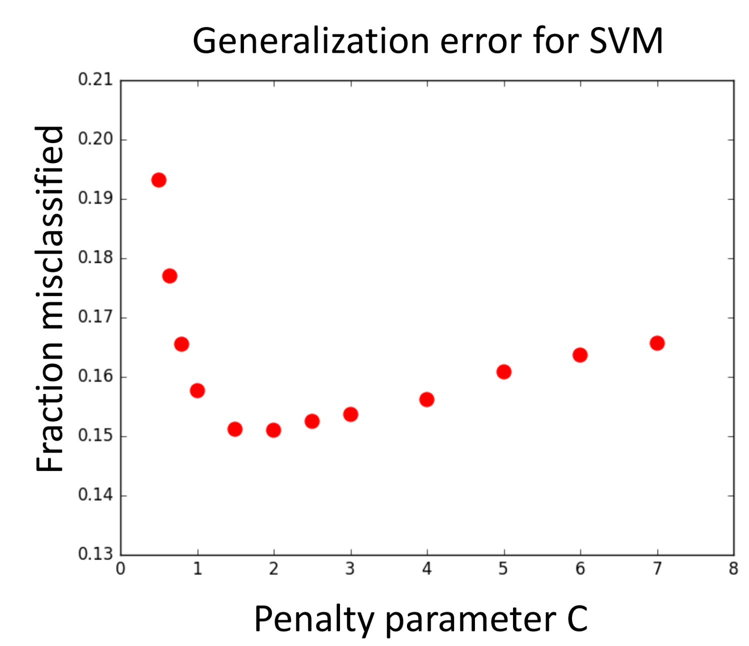


Figure 5. Generalization error for SVM with processed data for various values of C.

Generalization error of 15.3%, but hasn't leveled off yet.

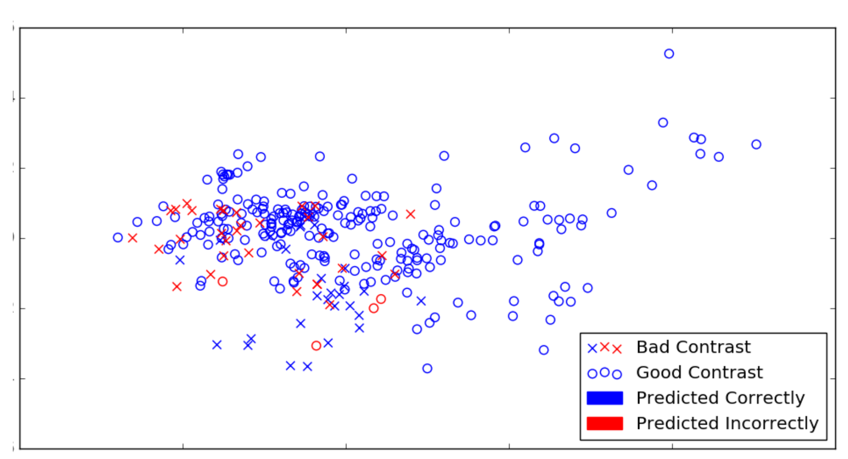


Figure 6. 2D projection onto principal axes of predictions by SVM. Red indicates misclassification.

True value \ Predicted value	0	1
0	0.44	0.56
1	0.03	0.97

Figure 7. Normalized confusion matrix for SVM

Can we predict the contrast (regression)?

Linear regression with or without powers of the features were not good models.

Weighted linear regression gives a flexible model which is analogous to SVM, error around 20% .

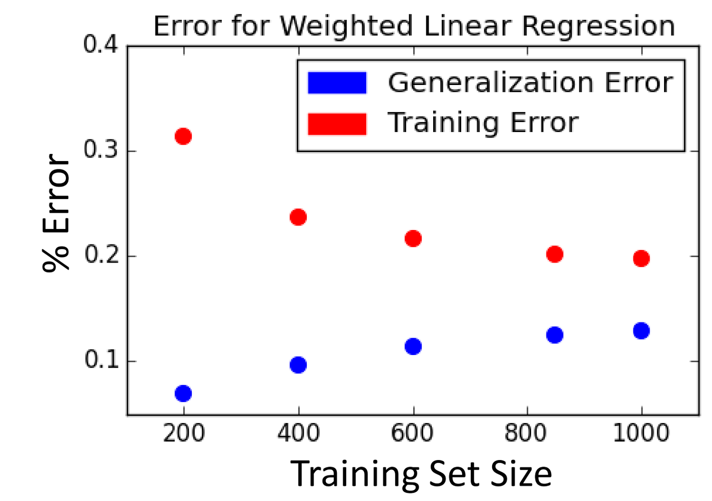


Figure 8. Generalization and training error for locally weighted linear regression for various training set sizes averaged 10 times.

Gaussian weights with γ=1.5

$$w_{xy} = \exp(-\gamma\|x - y\|_2)$$

Changed into classification problem to compare to SVM

True value \ Predicted value	0	1
0	0.63	0.37
1	0.11	0.89

Figure 9. Normalized confusion matrix for weighted linear regression.

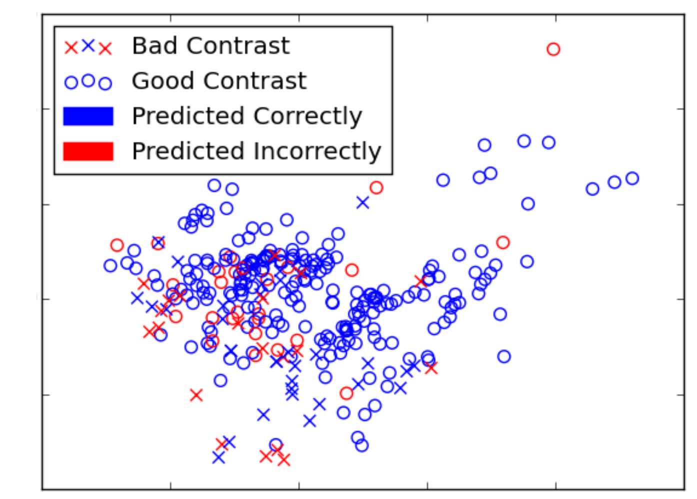


Figure 10. 2D projection of predictions by weighted linear regression. Red indicates misclassification.

Conclusion

SVM gives lower error and can be improved with better processing and choice of data. Currently limited by small amount of training data.

Weighted linear regression has higher error, but less misclassification of "bad" contrast data due to more information.

SVM with condensed data set and less features seems most promising for future work.

Acknowledgements

We thank Abhijith Rajan, Vanessa Bailey, Dmitry Savransky, Bruce Macintosh, and the rest of the GPI team for the data, preliminary work, and other advice.