Migraine Dataset

Mike Ghesquiere

CS8960

Dataset

- 484 data points.
- 27 Features.

  Date – 2012-01-01 to 2013-05-09. 5 days missing.

  Headache severity ∈ 0 (no migraine), 1 (mild), 2 (severe).

  Notes – Ignored.

  Triggers ∈ 1 (not present), 2, 3, 4 (significant presence)

  User selected, subjective, and trigger dependent.

Food triggers

Nightshade vegetables (tomatoes, eggplants, potatoes, peppers),

Smoked or cured meat (e.g., hot dogs),

Bananas

Aged or blue cheese

Chocolate

Sugar and Sweets

Drink triggers

Caffeine

Citrus fruit or juice

Beer

Red wine

Liquor or spirits

Ambient triggers

Cold air exposure

Perfume or strong odors

Bright or flashing Lights

Loud sounds

Changing weather

Hot and humid weather

Physical triggers

Physical exertion

Overslept

Lack of sleep

Post-stress letdown

Stress

Missed a meal

Dehydration

Additional constraints

Journal is recorded post-migraine (typically before bed).

Today's migraine may have been triggered yesterday.

Day of the week and time of year might be triggers.

Real-world data.

More interested in a diagnosis than a model.

Classification

Logistic regression from assignment 1:

$$>\mathbf{75\%} \text{ (depending on test/train split)}$$

The zero function (i.e. $f: X \to \{0\}$) has 68% accuracy.
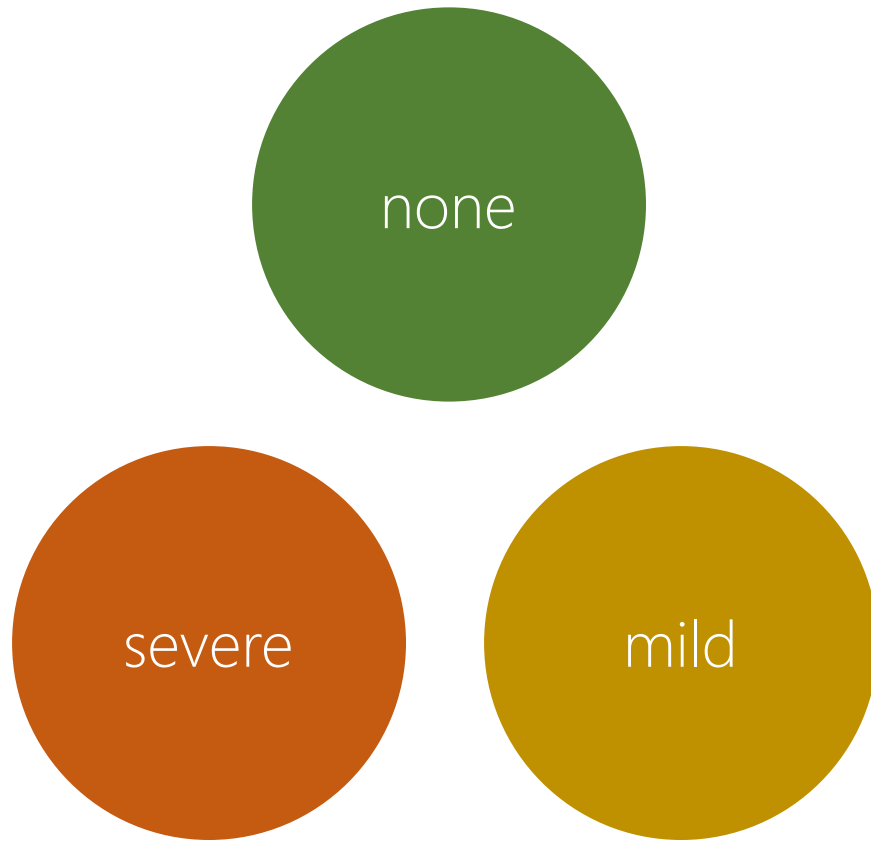
Classification

51 days with severe migraines
102 days with mild migraines
331 migraine-free days

Logistic Regresssion
Three class, "One-vs.-All"
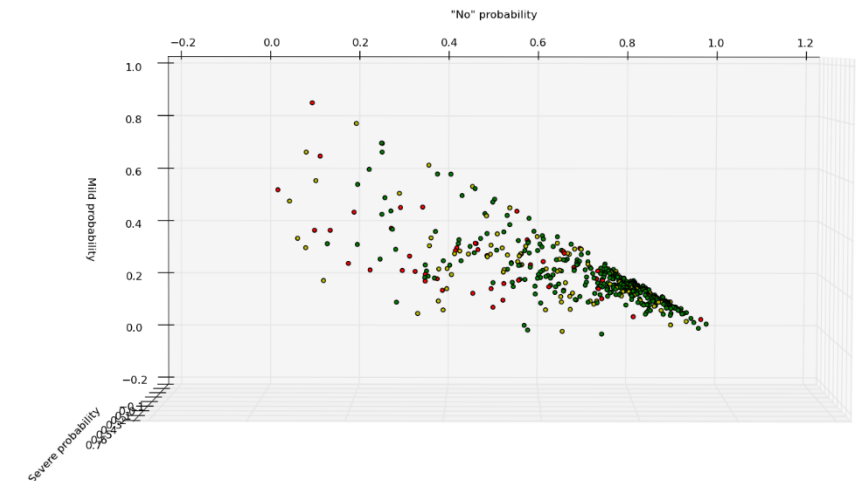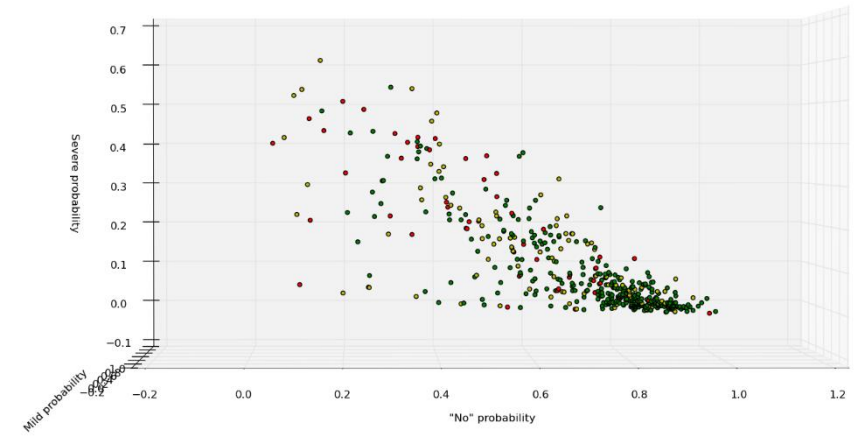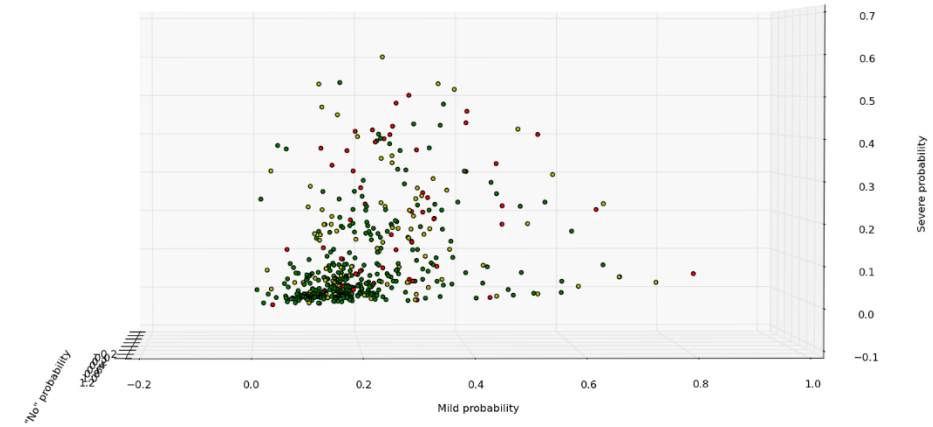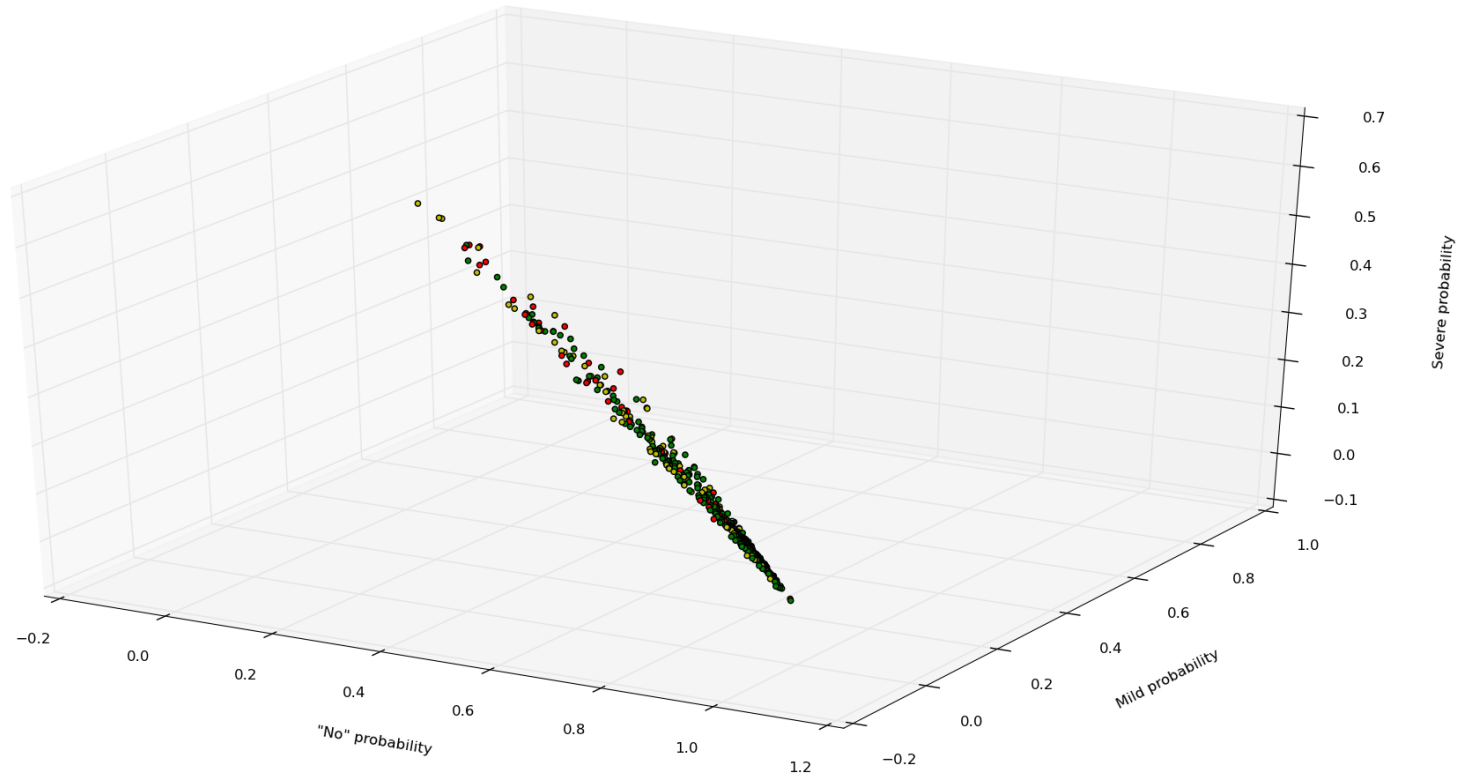
Classification

Classification
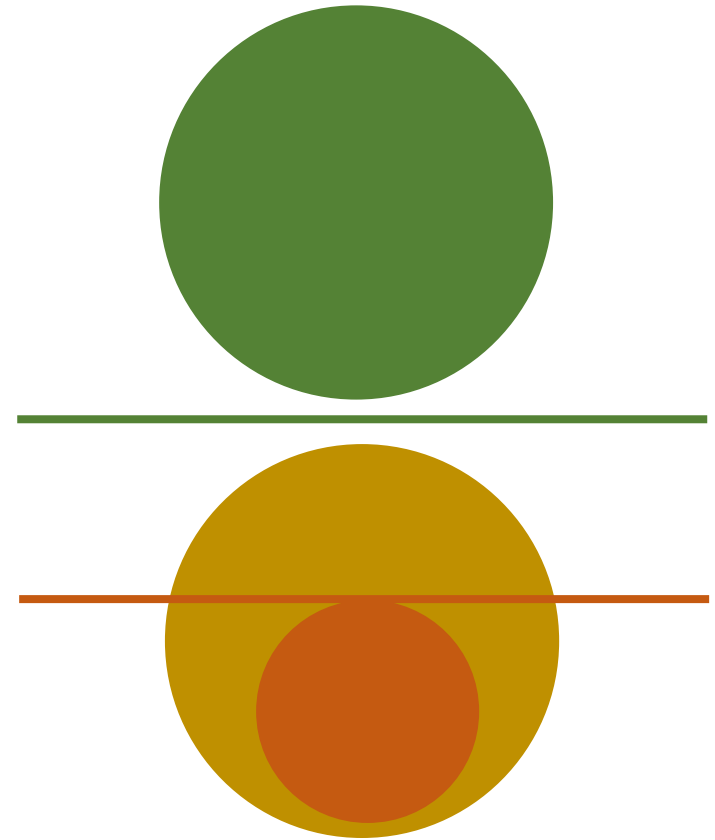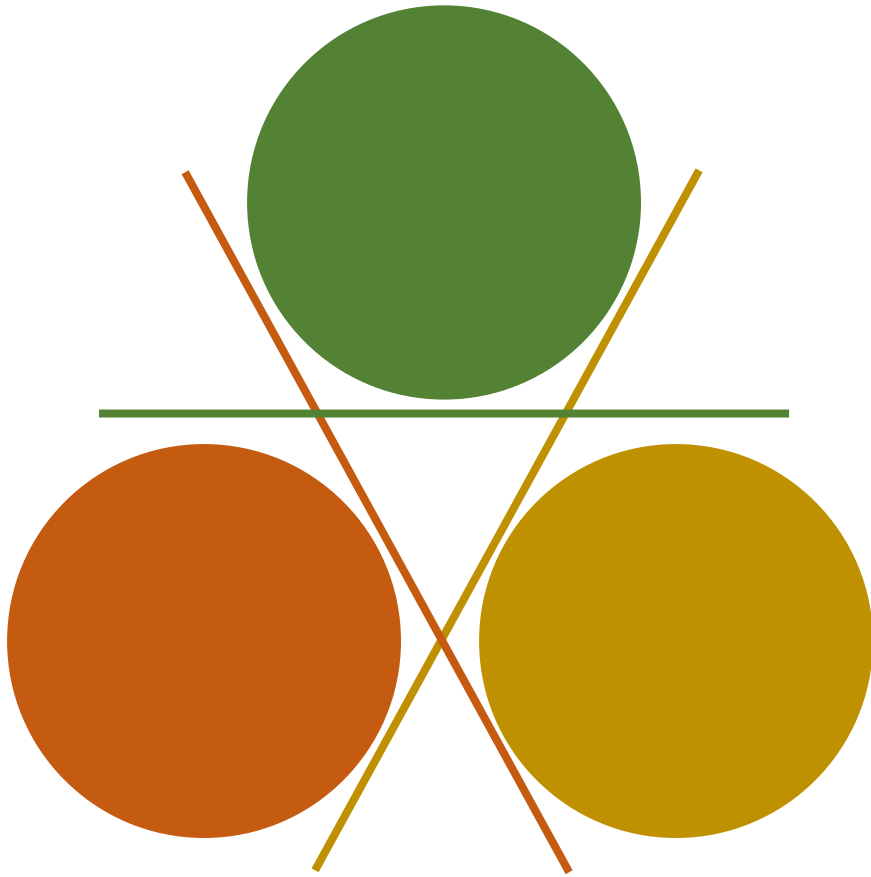
Classification

$$[0] \quad [1 \qquad 2] \quad = clf^{(1)}$$

$$[0 \qquad 1] \quad [2] \quad = clf^{(2)}$$

$$[0] \quad [1] \quad [2] \quad = clf^{(3)}$$

$$\approx clf^{(1)} + clf^{(2)} \ ?$$

Comparing Classifiers: Oracle

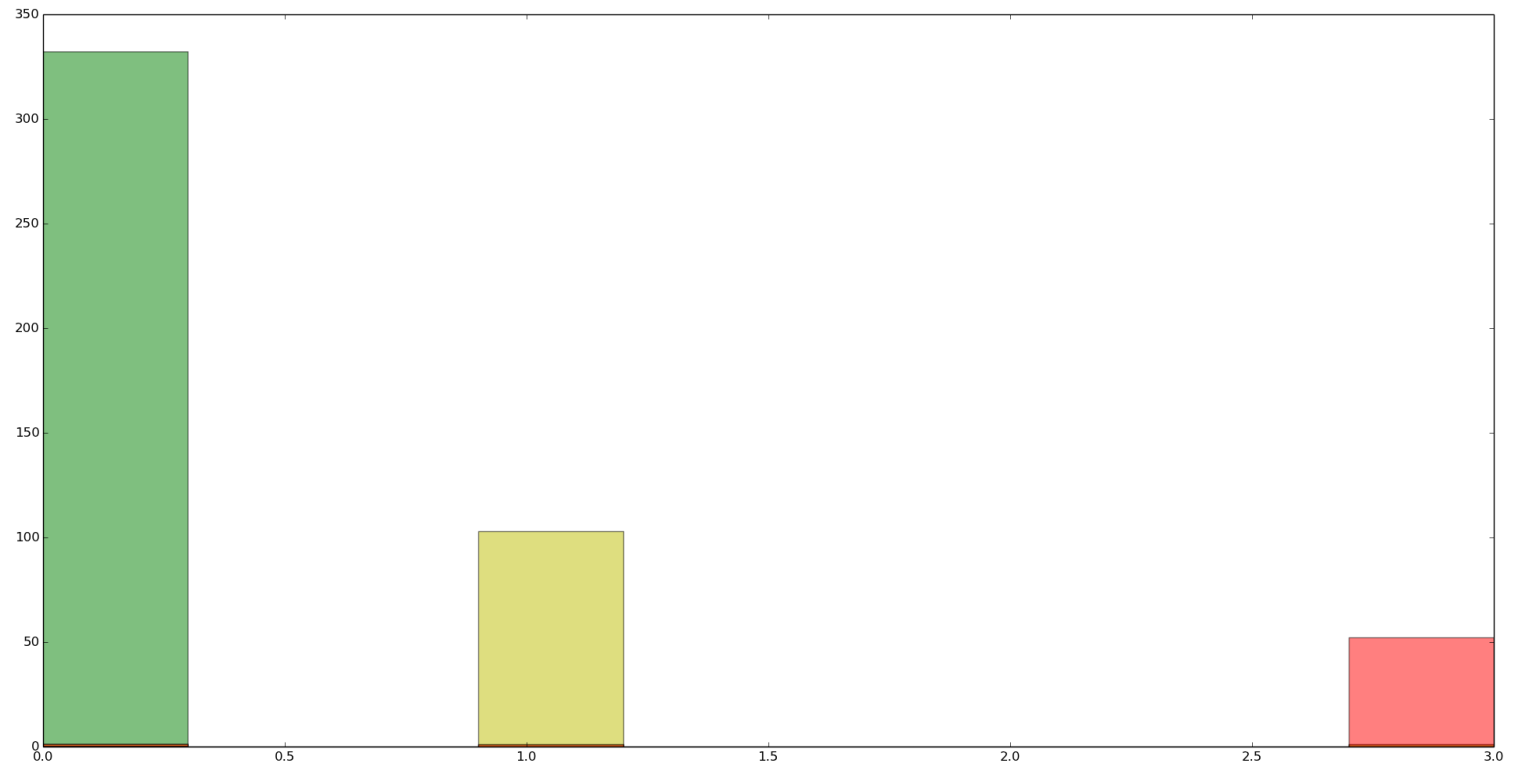Define:
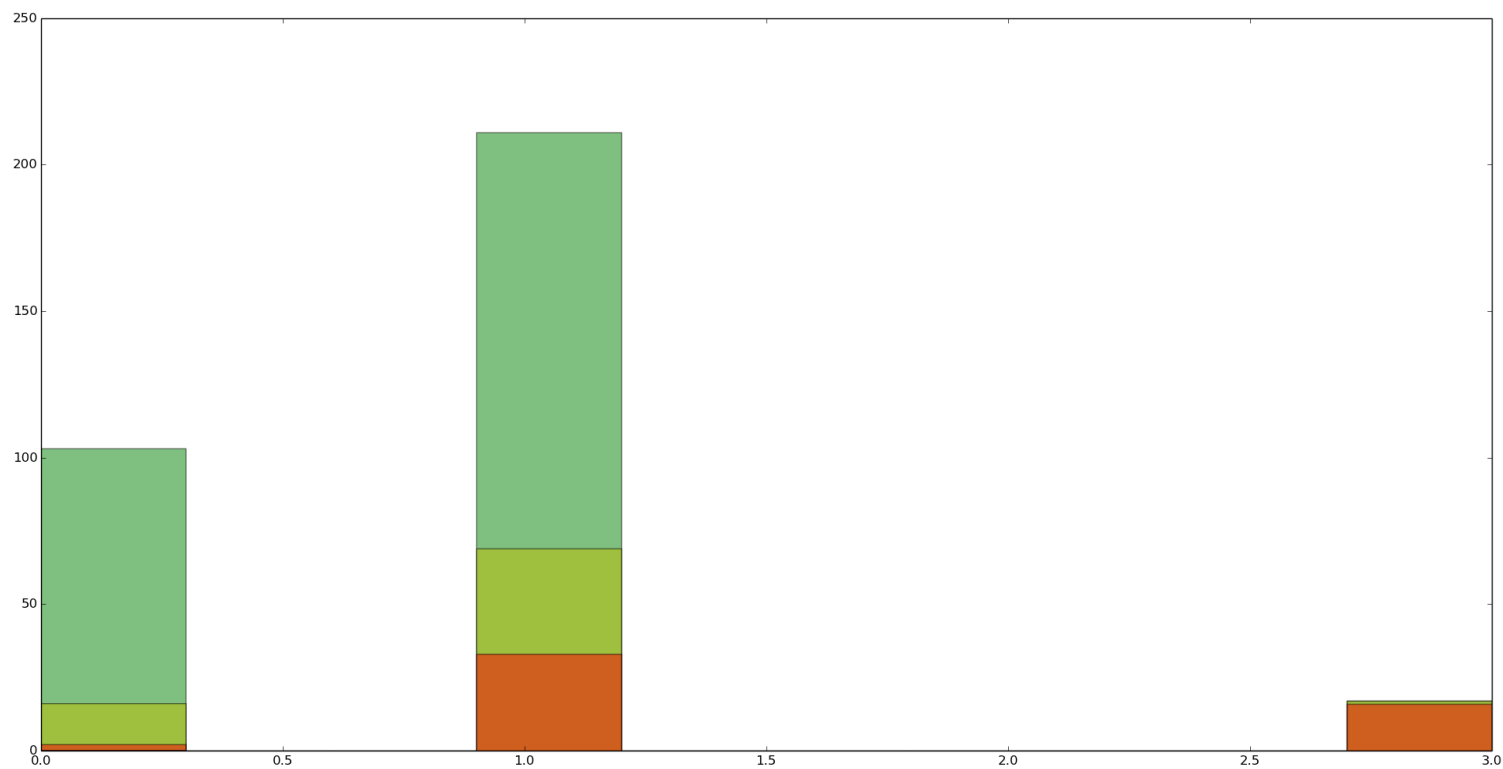$$clf^{(4)} := clf^{(1)} + \mathbf{2}\, clf^{(2)}$$

Thus:
$$2 \notin Im\big(clf^{(4)}\big)$$
$$\Longleftrightarrow$$
$$\big(clf^{(1)}(x) = 1\big) \Longrightarrow \big(clf^{(2)}(x) = 1\big)$$

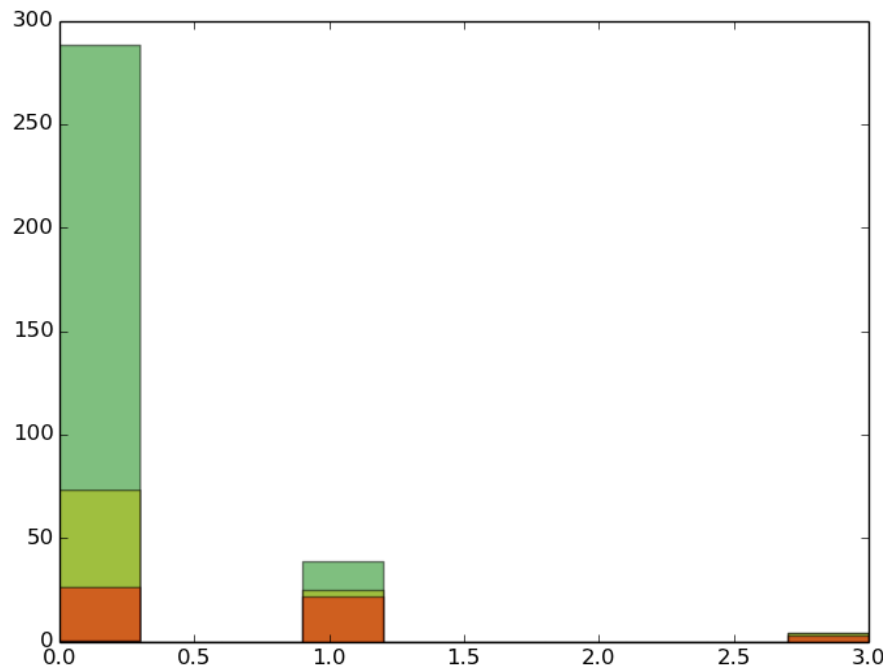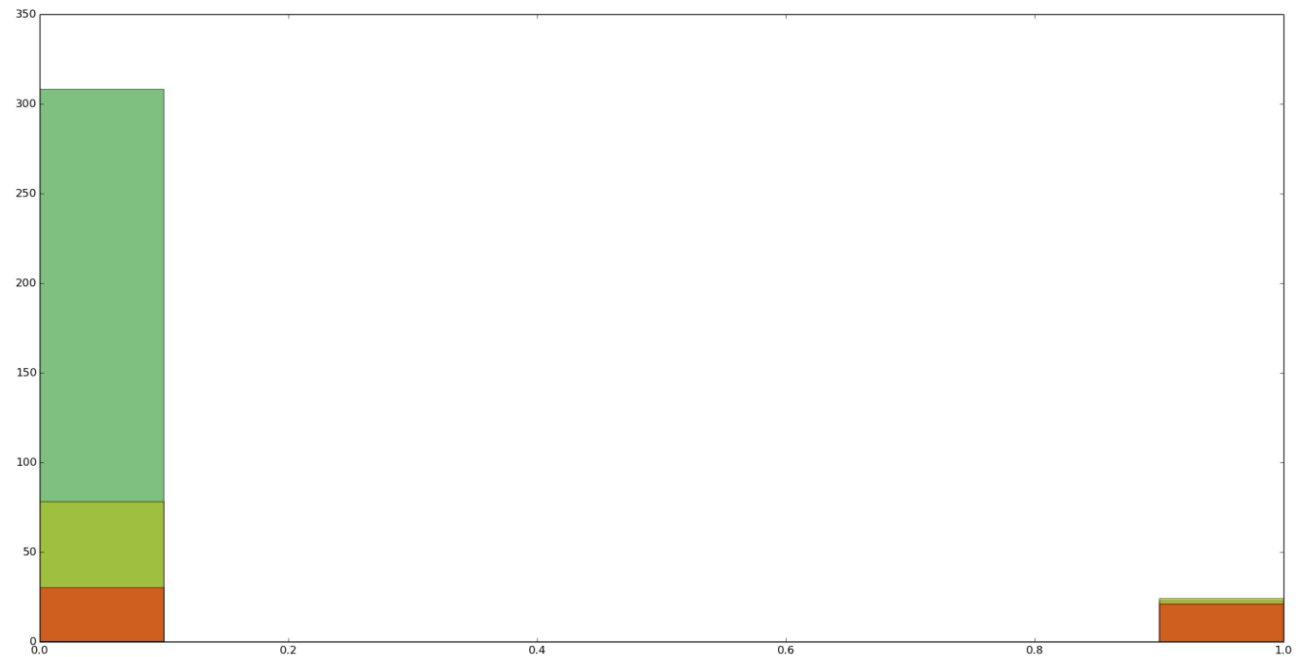*Ensures that $\mathbf{clf^{(2)}}$ is more strict than $\mathbf{clf^{(1)}}$.*

Logistic Regression

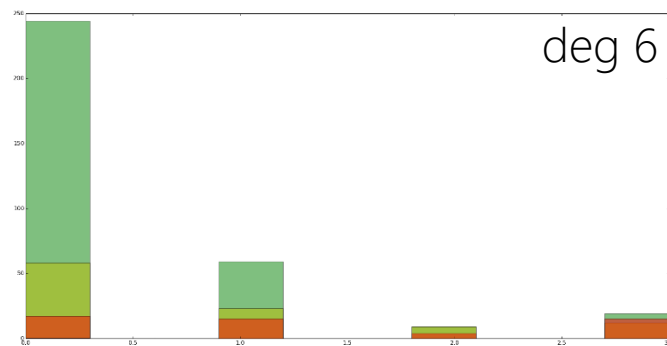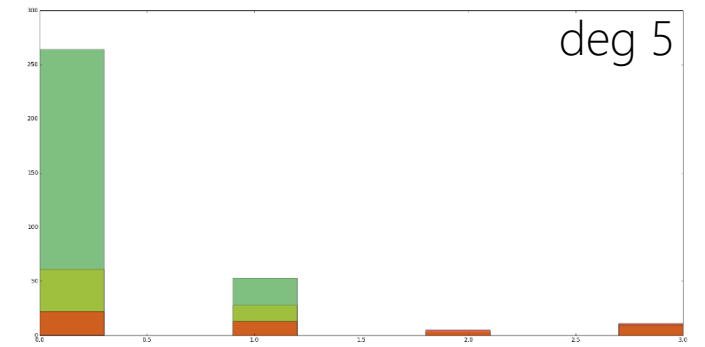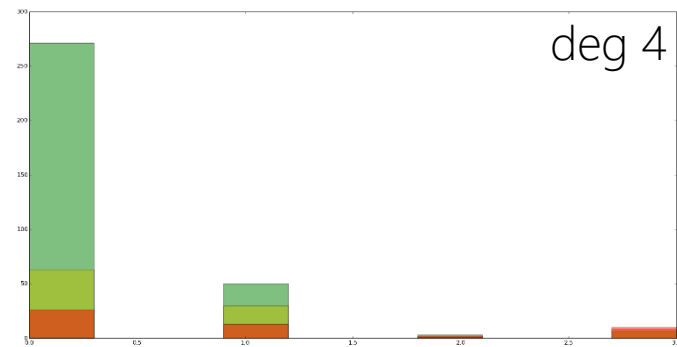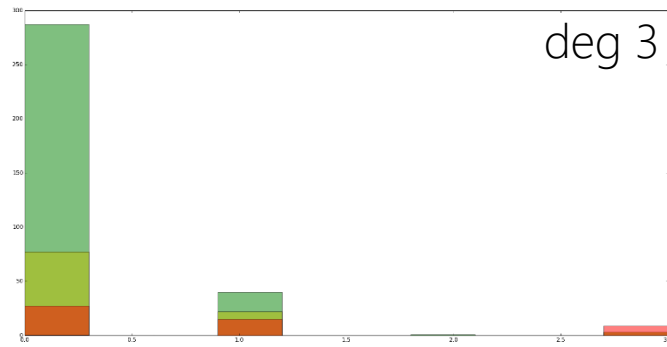# Support Vector Classifier: Linear/rbf kernal

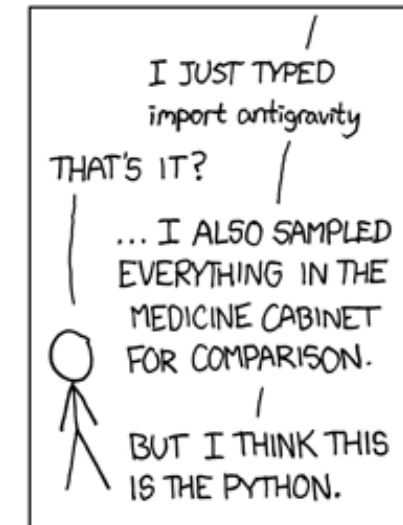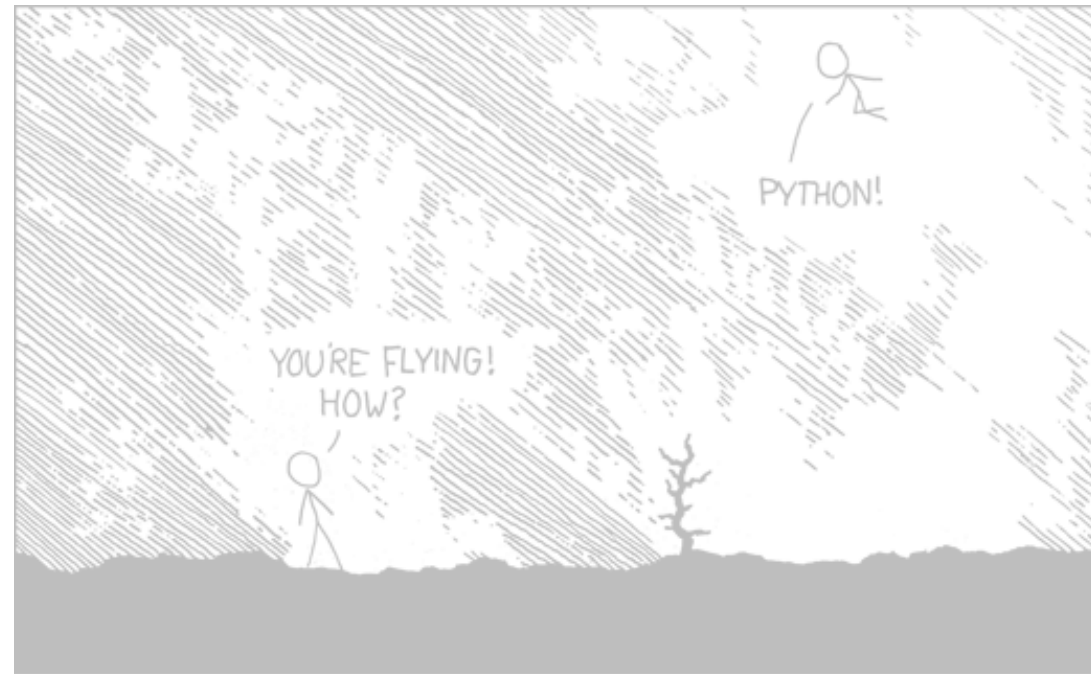## Linear kernal

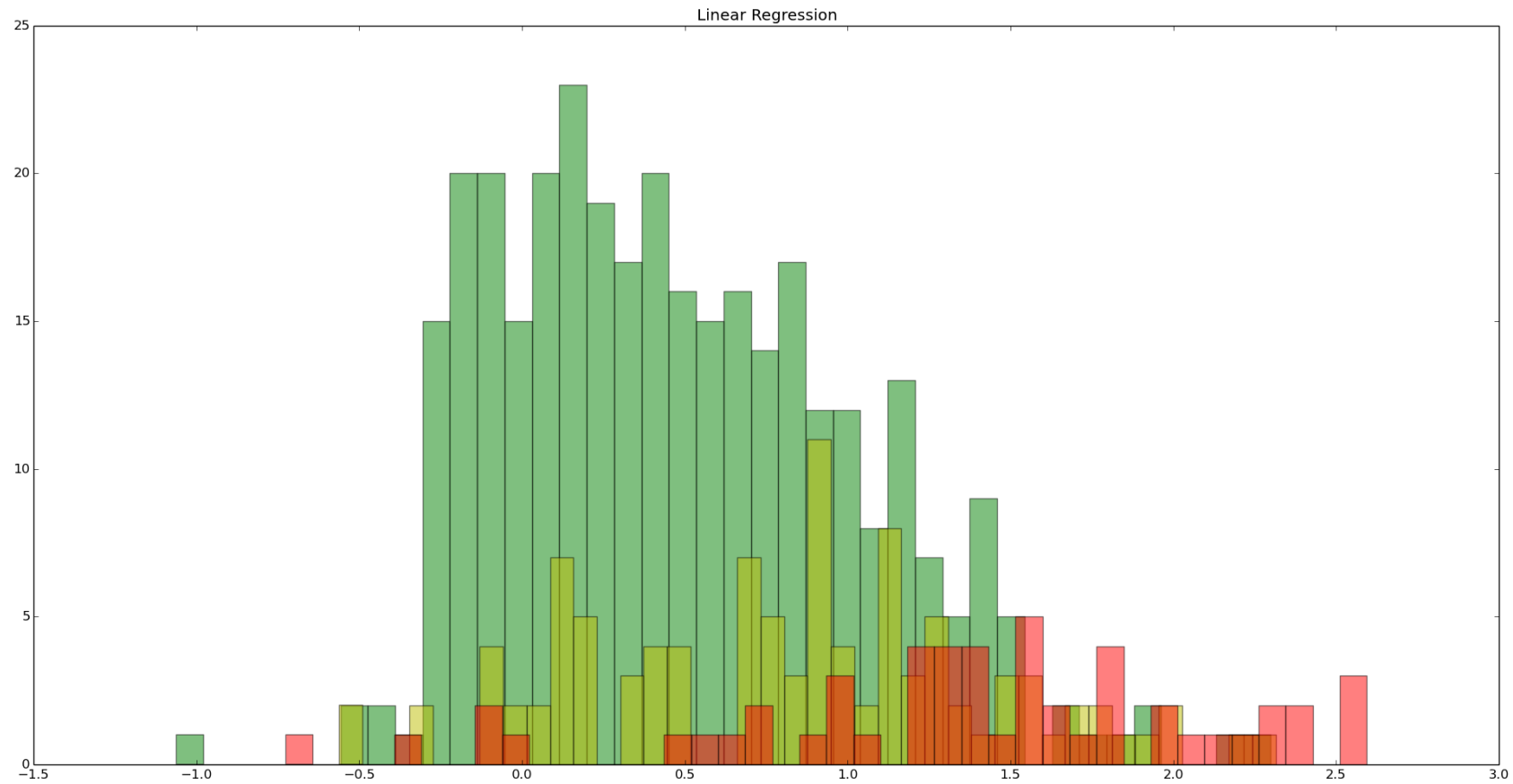## rbf kernal

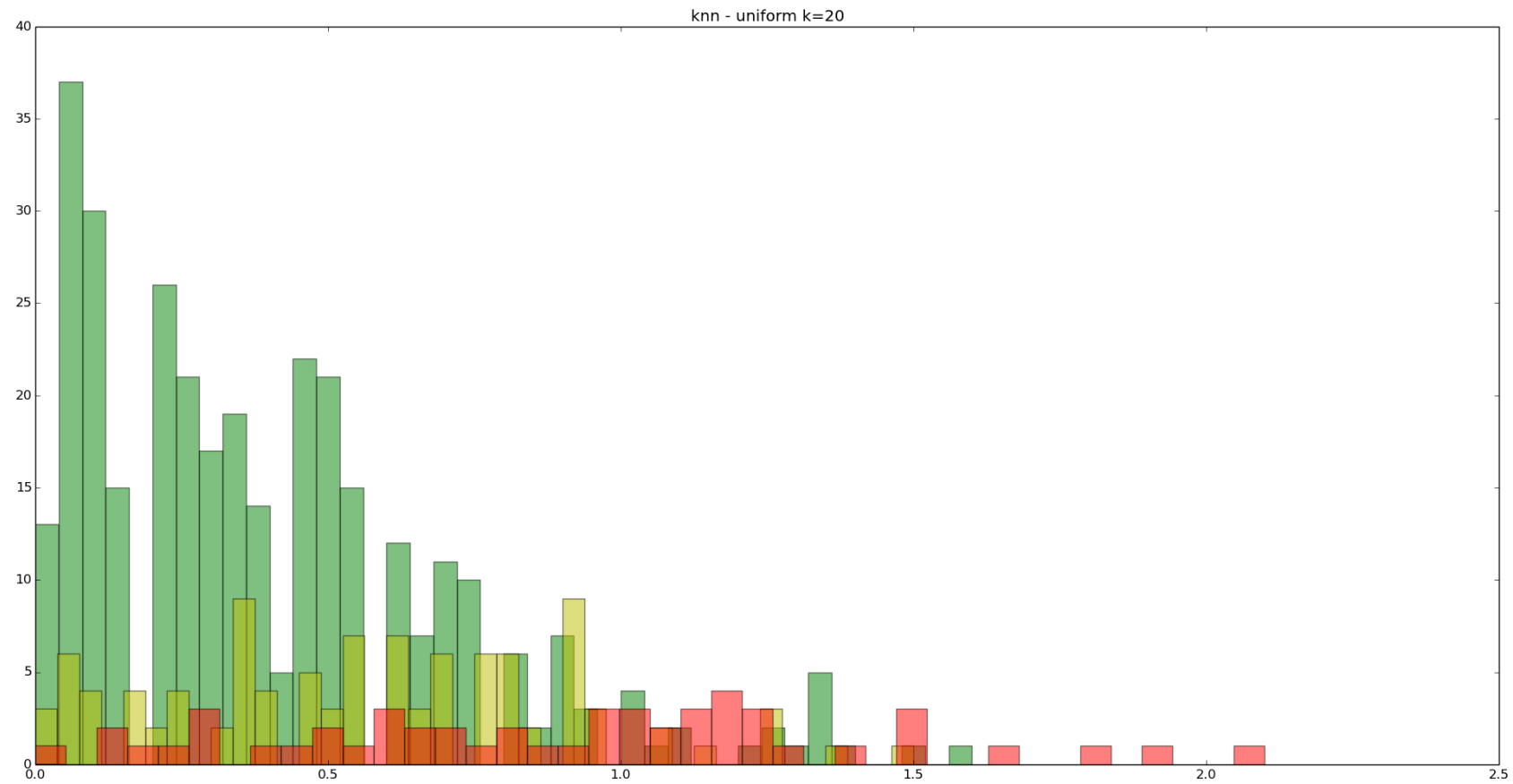# Support Vector Classifier: Polynomial kernal
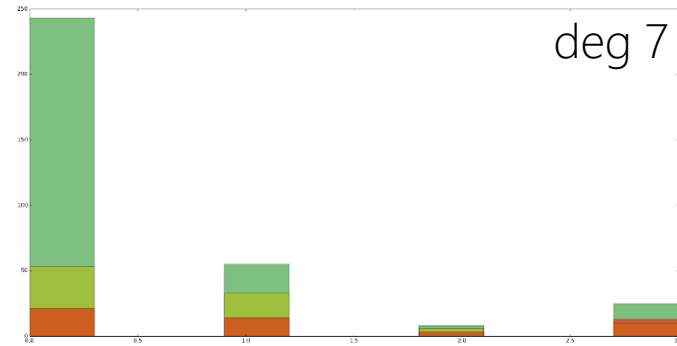
Regressor Selection

Python Medicine

Cabinet algorithm

Linear Regression
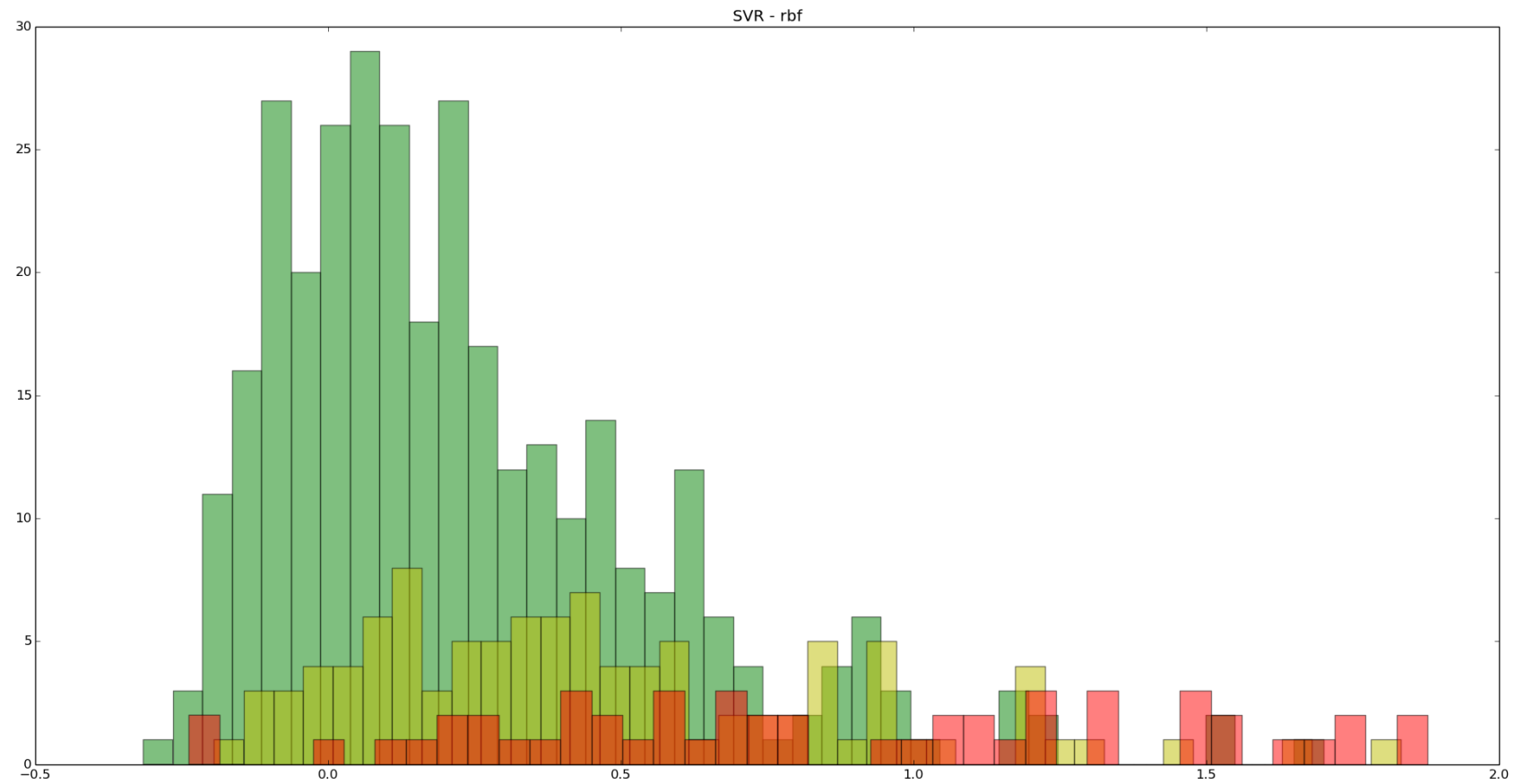
k-NN (k=20)

knn - uniform k=20

Polynomial SVR (degree 3)
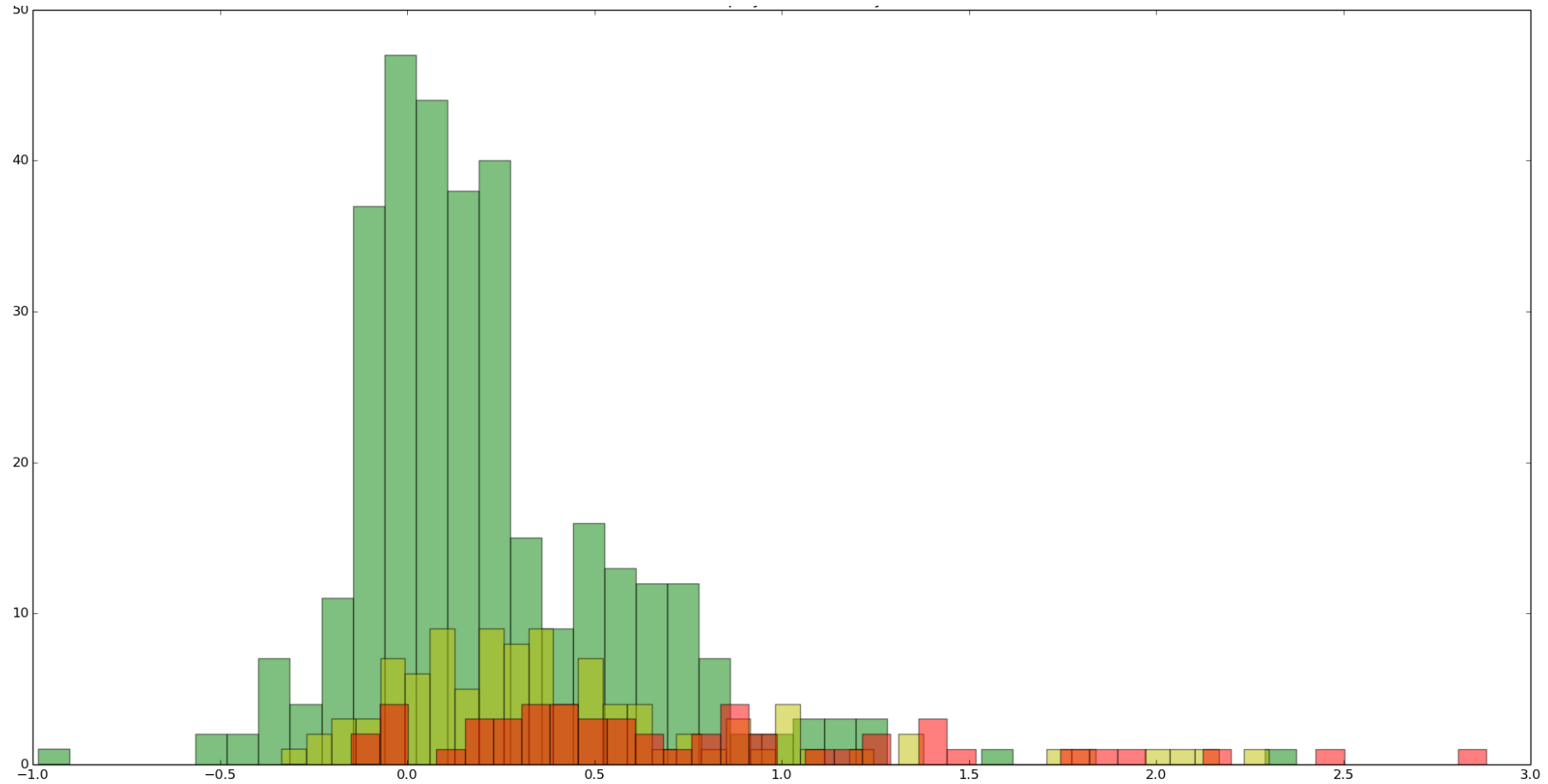
rbf SVR

SVR - rbf

Regression

Linear Regression (normalized, unnormalized)

kNN ($k \in \mathbf{1,2,5,10,15,20,30,40,50,75,100}$)

SVR (linear, rbf, poly w/ $\mathbf{deg \in \{3,4,5,6,7,8,9\}}$)

Ridge Regression ($\alpha \in \{0.1, 0.2, \dots 0.9\}$)

Lasso, LARS Lasso

Orthogonal Matching Pursuit

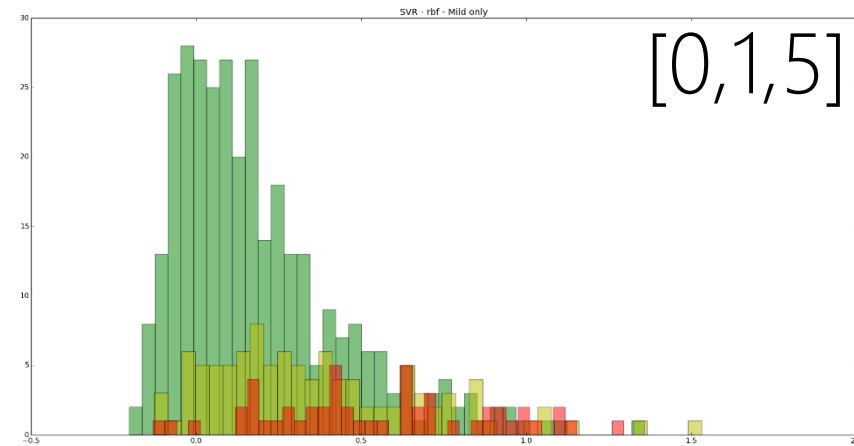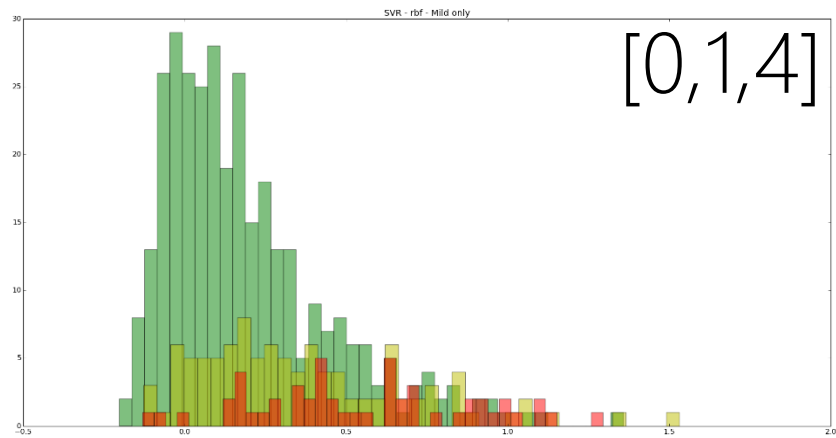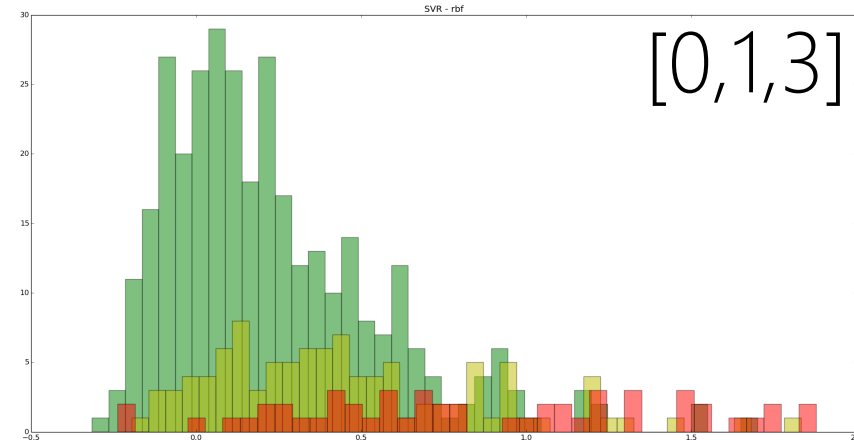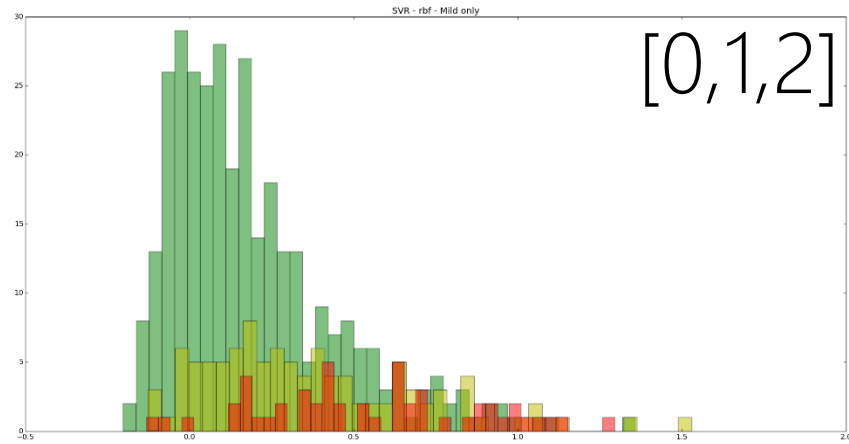Bayesian Regression (Ridge, ARD)

Decision Tree

Changing the Scale

"0", "1" and "2" are class labels, not magnitudes.

What happens when we vary their relative weight?

Only need to change "2".

Changing the Scale

Further Work

Feature Selection.

Python Medicine Cabinet

Subsets of Triggers

predictable e.g. weather

changeable e.g. diet

More days.

Incorporating "notes"

Effects of medication. Tylenol vs. Zomig vs. Advil?

Thank you.