Quantifying Latent Fingerprint Quality

Team: Jordan Varney, Sarah Scheffler (Fall PM), Christopher Eriksen, Martin Loncaric (Spring PM),

Adviser: Yi-Chieh (Jessica) Wu

Liaison: Nicholas Orlans and Sarah Doyle

Harvey Mudd Clinic

HARVEY
MUDD
COLLEGE

Project Goal

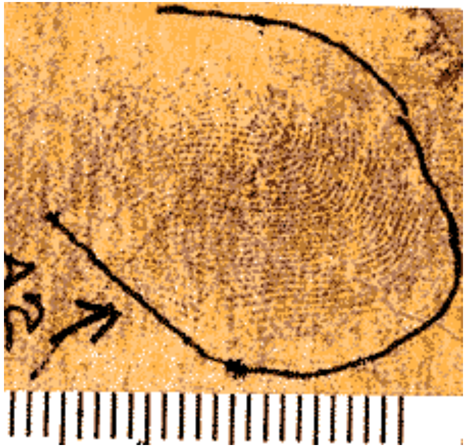
Develop and implement a mathematical model for **latent** fingerprint **quality** with regard to **AFIS matching** and assess the performance of various **quality features**.

Fingerprint Overview

Exemplar Prints

- Taken on purpose
- Comprise databases

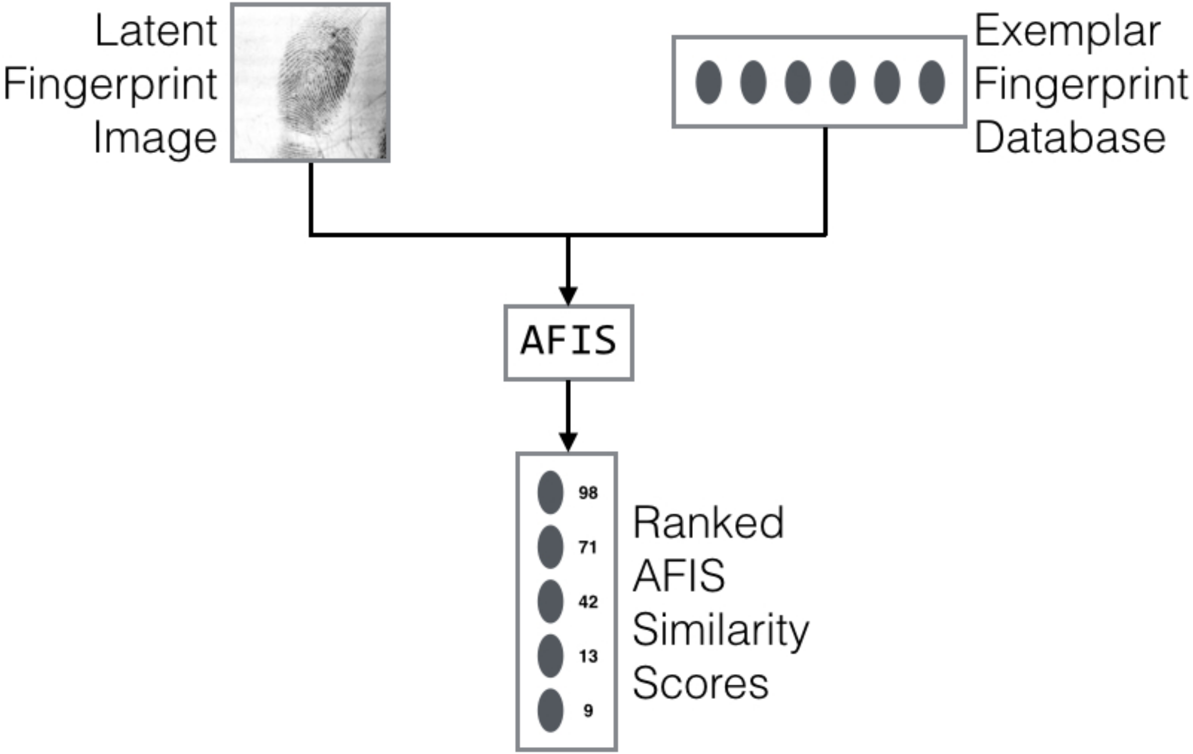
Latent Prints



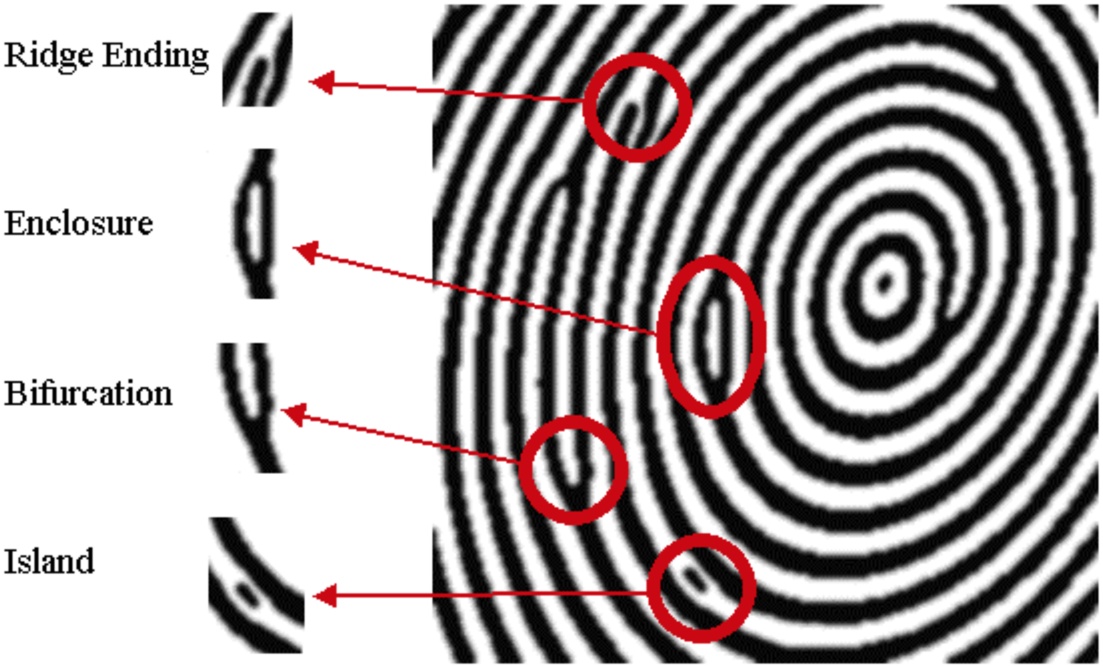
- Crime scene prints
  - Incomplete
  - Background noise
  - Unknown orientation

www.vetmed.vt.edu,
wilenet.org

Automated Fingerprint Identification System



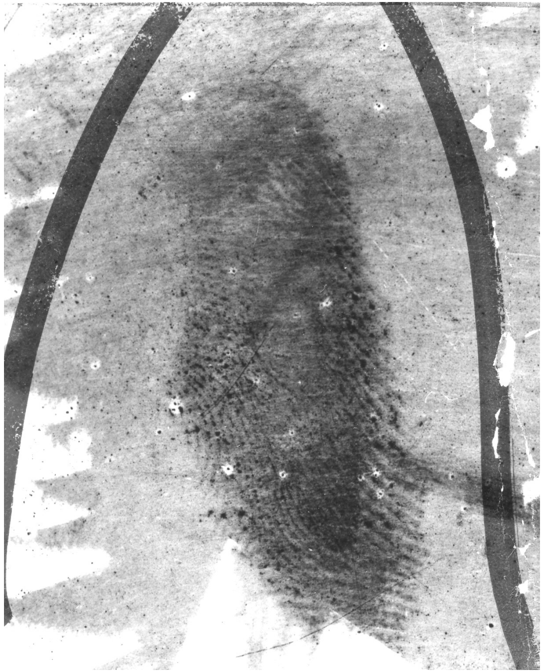
Minutiae



Latent Suitability for AFIS Identification

“Good”

“Bad”



NIST Special
Database 27A

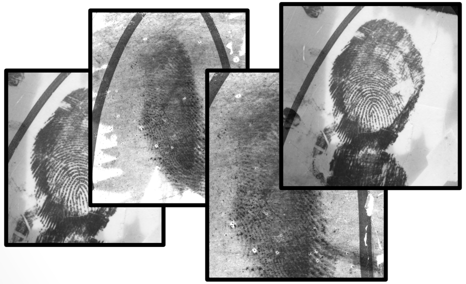
Project Goal

Develop and implement a mathematical model for **latent** fingerprint **quality** with regard to **AFIS matching** and assess the performance of various **quality features**.

Quality Scores

Problem:

- Some latents are not suitable for AFIS identification
- Too many prints, not enough AFIS time

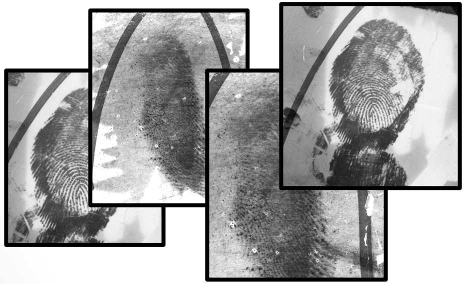


NIST Special Database 27A

Quality Scores

Problem:

- Some latents are not suitable for AFIS identification
- Too many prints, not enough AFIS time

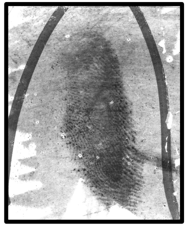


NIST Special Database 27A

Solution:

- Model latent fingerprint image quality
- Only use AFIS for good quality latents

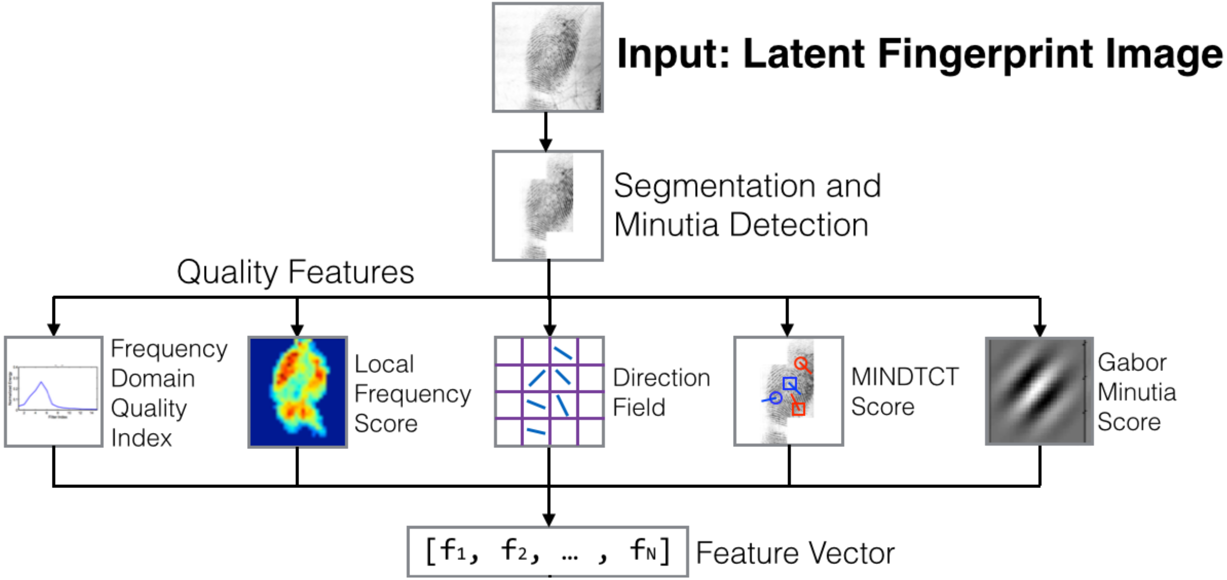
>



Project Goals

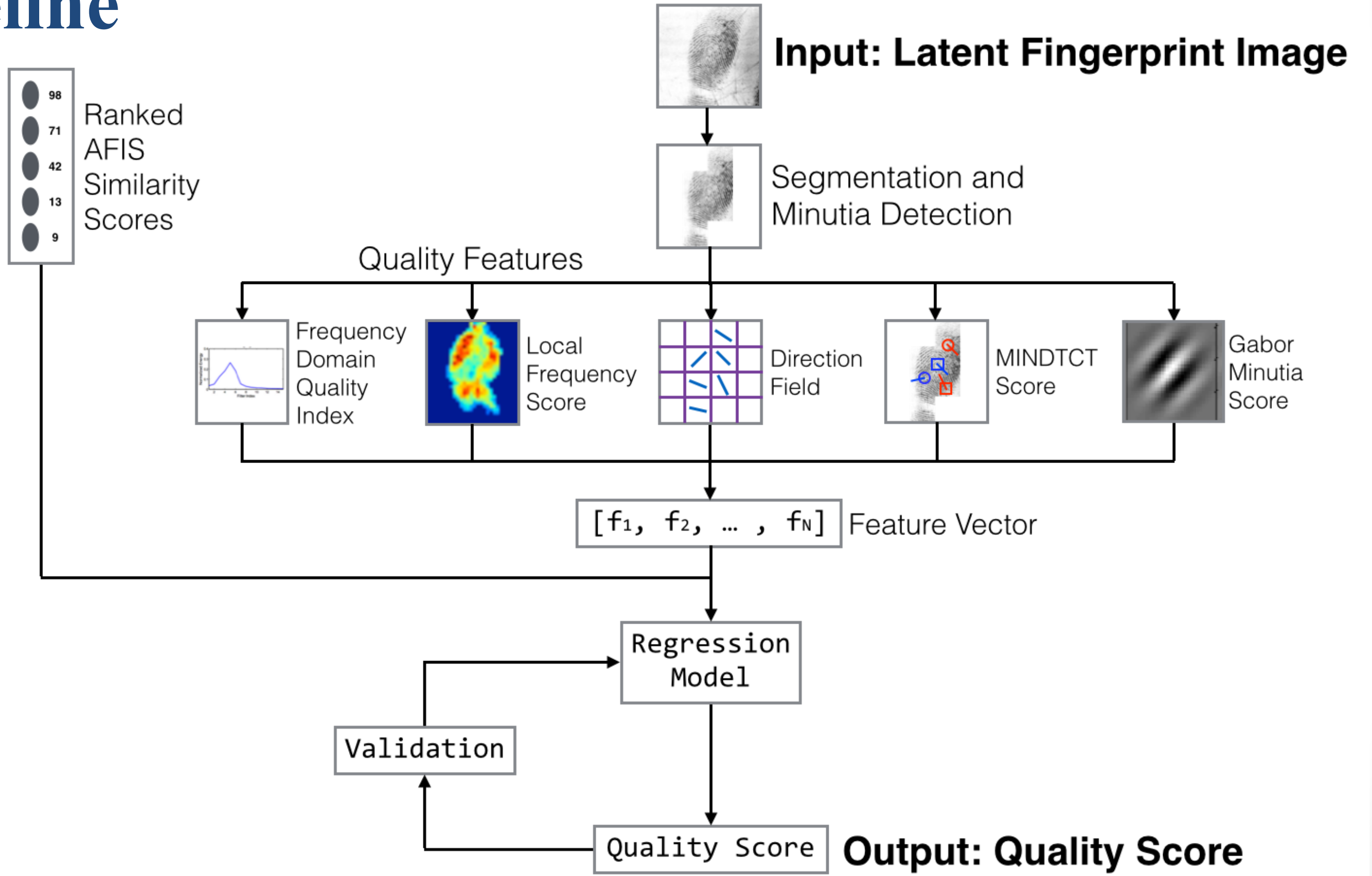
- Assess latents' suitability for identification with AFIS
- Analysis of existing fingerprint quality metrics
- Mathematical model for latent fingerprint quality
- Implementation of quality score

Pipeline



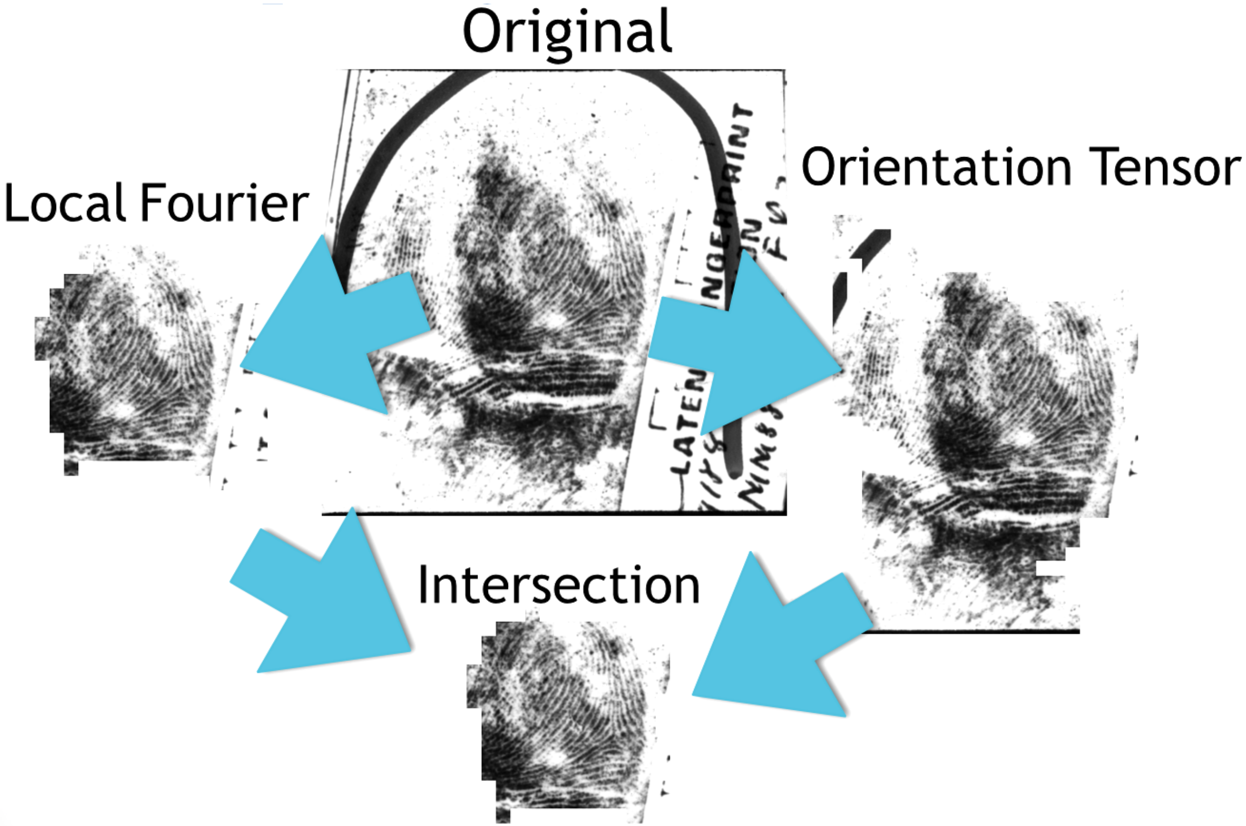
Input: Latent Fingerprint Image
Segmentation and Minutia Detection
Quality Features
Frequency Domain Quality Index
Local Frequency Score
Direction Field
MINDTCT Score
Gabor Minutia Score
[f1, f2, … , fN]
Feature Vector

Pipeline



Input: Latent Fingerprint Image
Segmentation and Minutia Detection
Ranked AFIS Similarity Scores
98
71
42
13
9
Quality Features
Frequency Domain Quality Index
Local Frequency Score
Direction Field
MINDTCT Score
Gabor Minutia Score
[f1, f2, … , fN]
Feature Vector
Regression Model
Validation
Quality Score
Output: Quality Score

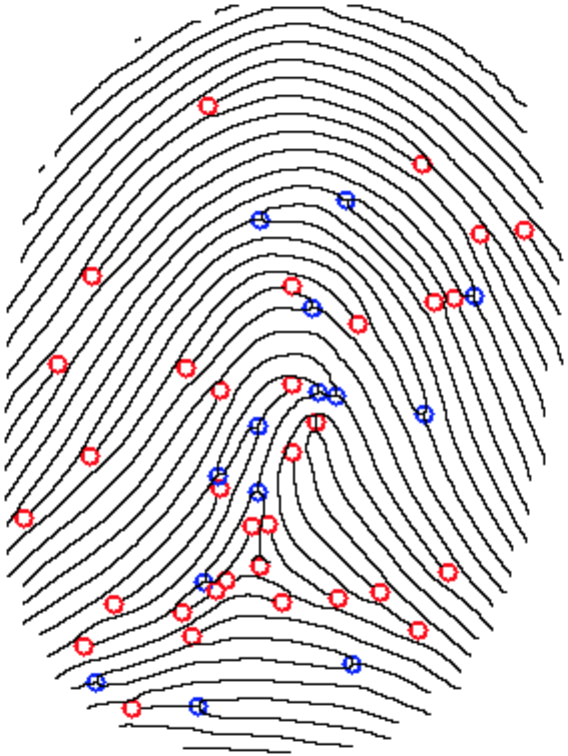
Segmentation



NIST Special
Database 27A

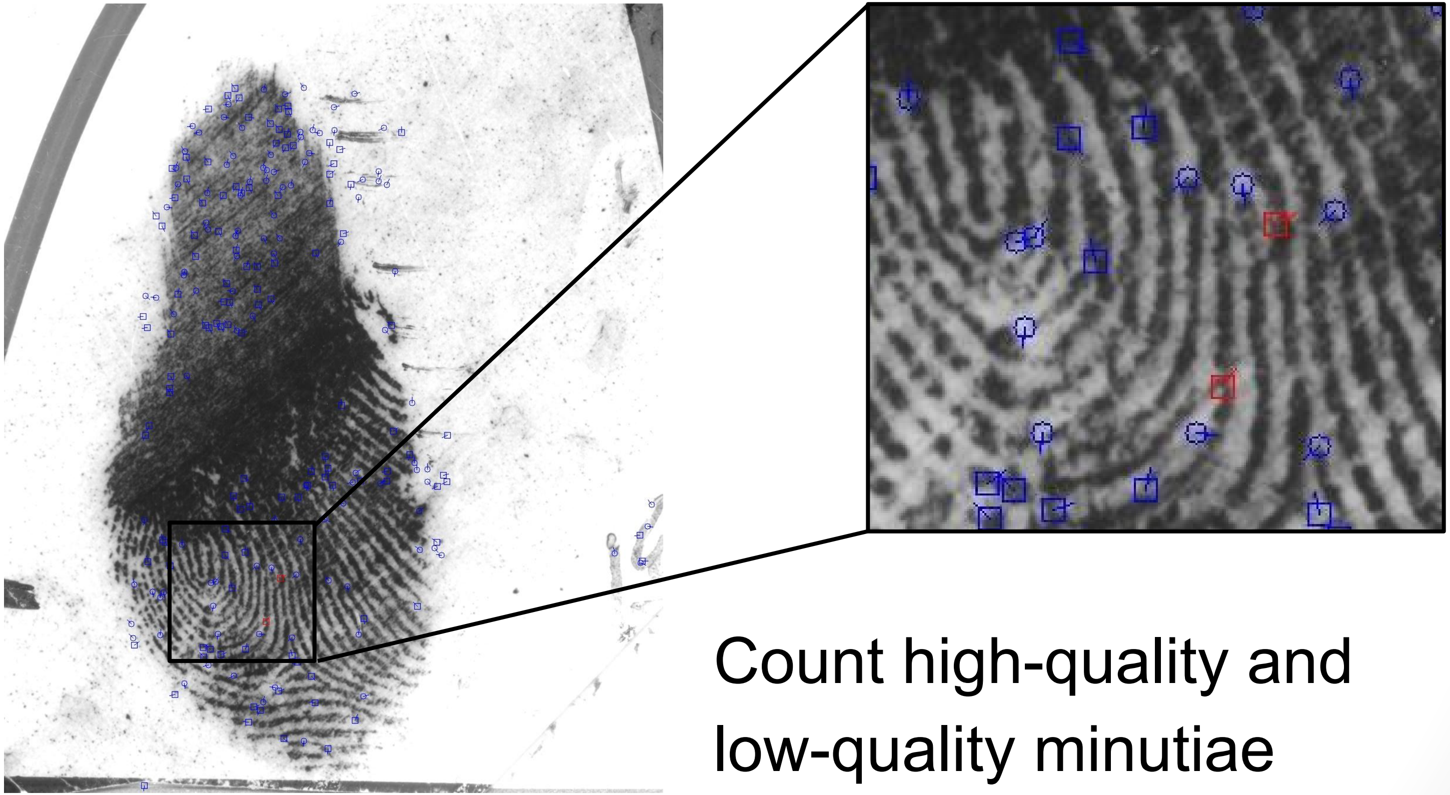
Minutia Detection

```
Image (w,h) 516 485

60 Minutiae Detected

   0 :  209,  117 : 13 :  0.058 :RIG
   1 :  212,  164 : 14 :  0.057 :RIG
   2 :  216,  188 : 14 :  0.122 :RIG
   3 :  218,  120 : 13 :  0.124 :RIG
   4 :  224,  151 : 30 :  0.123 :RIG
   5 :  224,  203 : 27 :  0.061 :BIF
   6 :  225,  159 : 11 :  0.123 :BIF
   7 :  240,  104 :  0 :  0.058 :BIF
   8 :  245,  156 :  5 :  0.125 :RIG
   9 :  246,  193 :  9 :  0.124 :BIF
  10 :  248,  164 :  9 :  0.128 :BIF
  11 :  248,  200 : 25 :  0.058 :BIF
  12 :  251,  134 :  2 :  0.124 :RIG
  13 :  255,  147 :  3 :  0.122 :BIF
  14 :  260,  126 :  5 :  0.126 :RIG
```



NIST Special Database 27A
NBIS MINDTCT
http://www.griaulebiometrics.com/

Good and Bad Minutia Counts



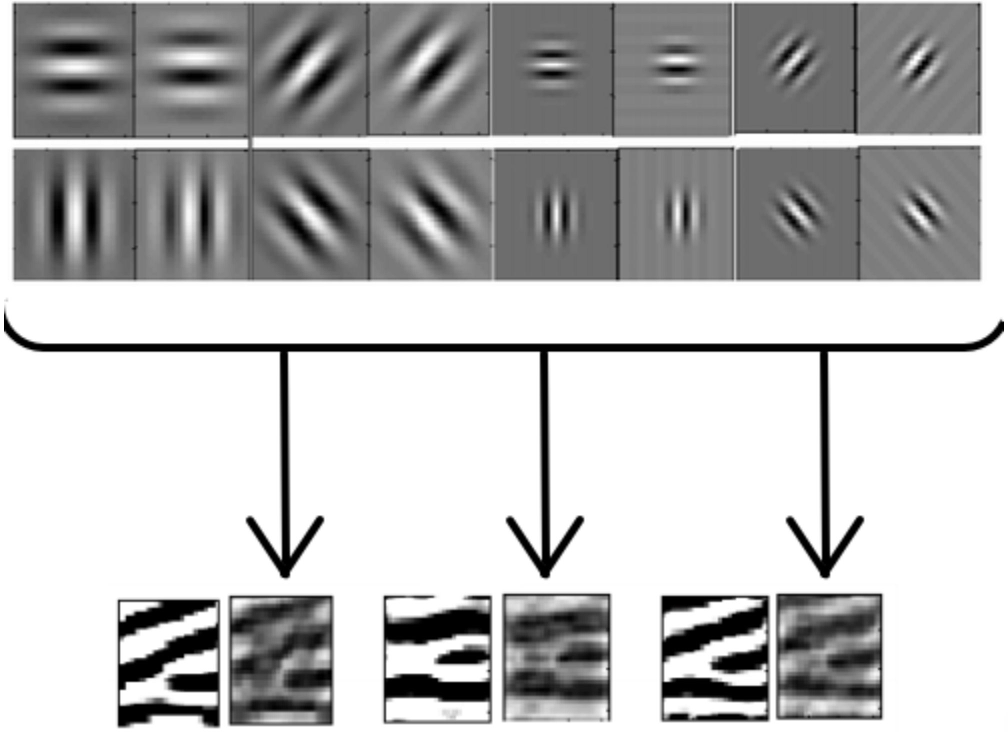
NIST Special Database 27A

Count high-quality and low-quality minutiae

Gabor Minutia Score

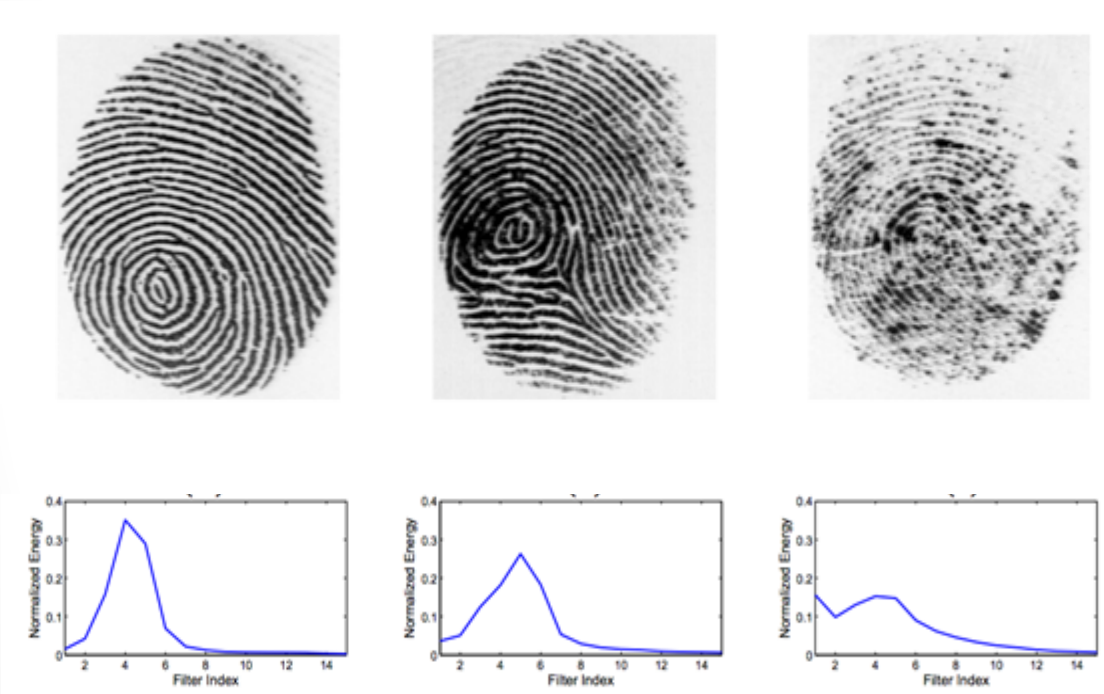
Recreate a minutia using basis functions

Learn mapping of coefficients to quality



Chikkerur et. al. 2005.

Frequency Domain Quality Index

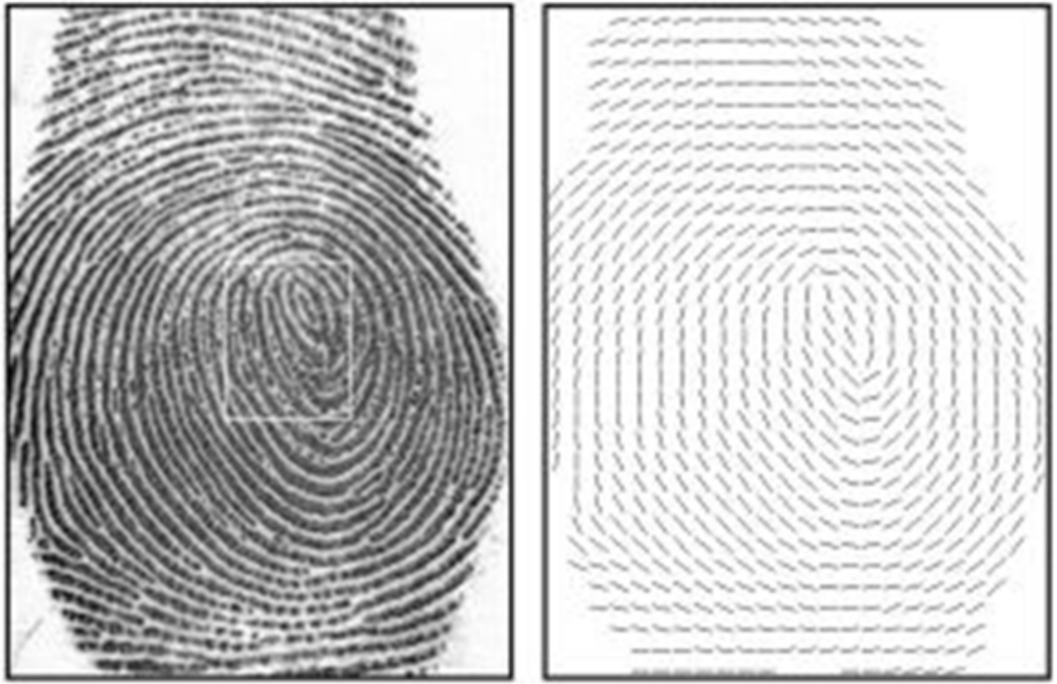


High-quality prints have narrow peaks in the frequency domain

Chen, Yi et. al. 2005.

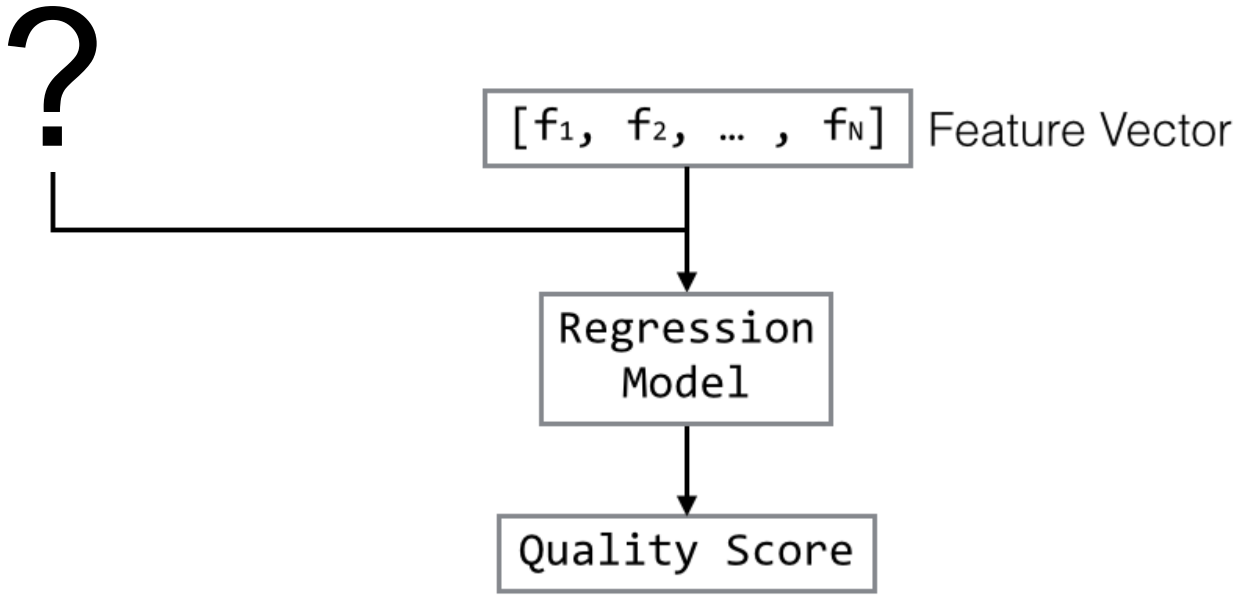
Direction Field

Measure ridge continuity

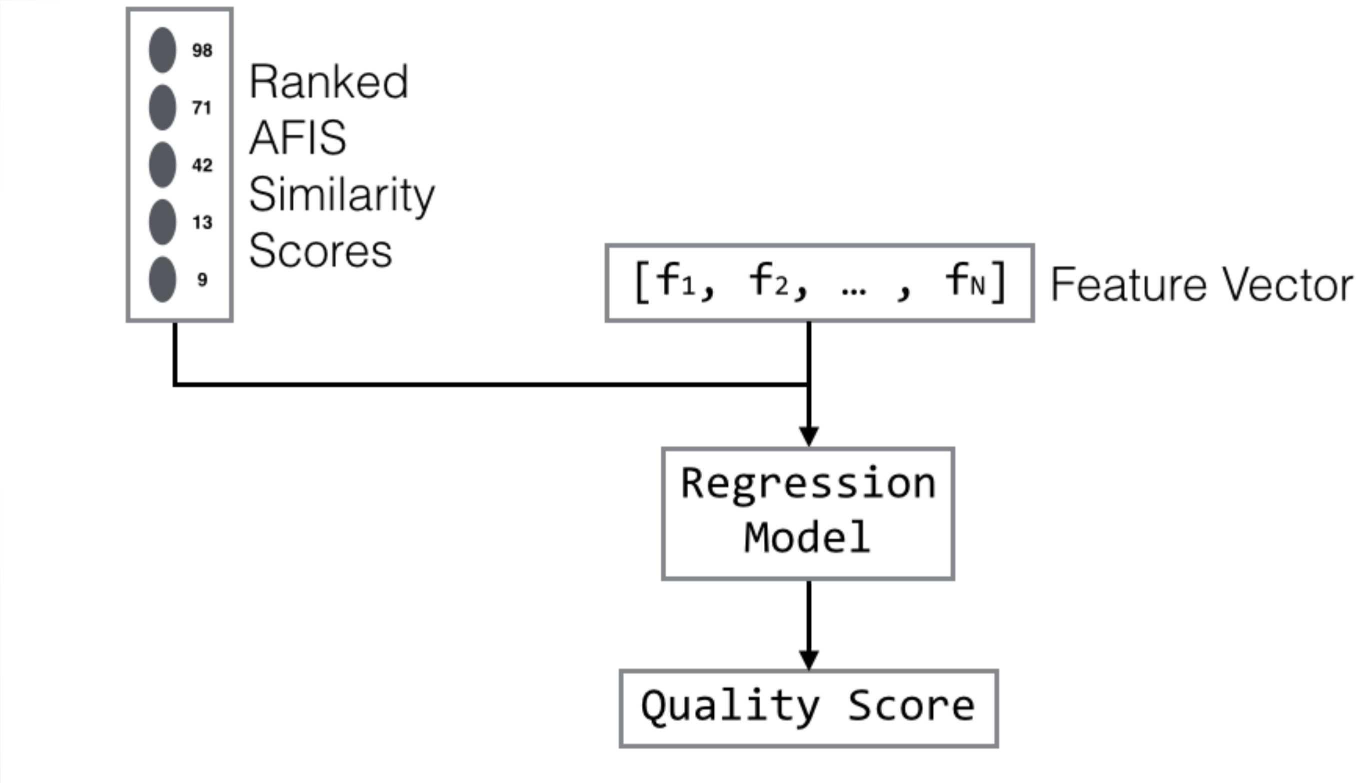


Zaeri, Naser. 2011.

Predicting Quality from Features



Predicting Quality from Features



Response Variable

- **Quality**: The chance that the correct match (assuming it exists) will appear in the top 20 ranked AFIS results
- **Response**: Our approximation of quality, derived from AFIS results

Response Variable

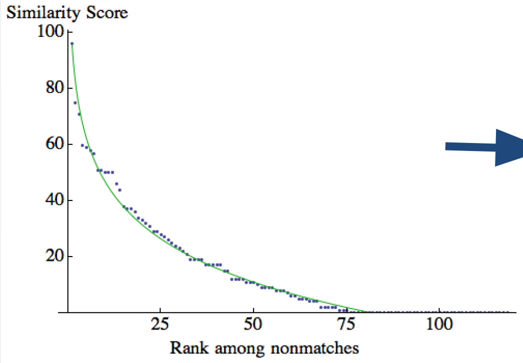
$$q = \int_0^\infty Q(s)R(s)ds$$

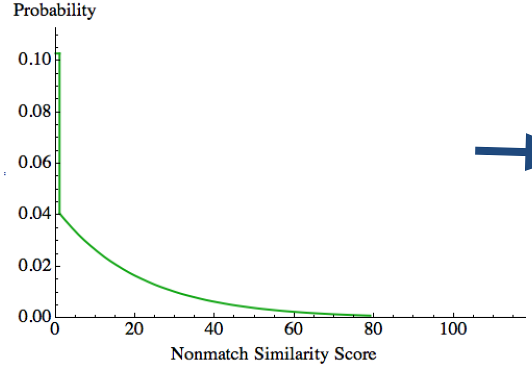
Q(*s*): probability that correct match will have similarity score *s*

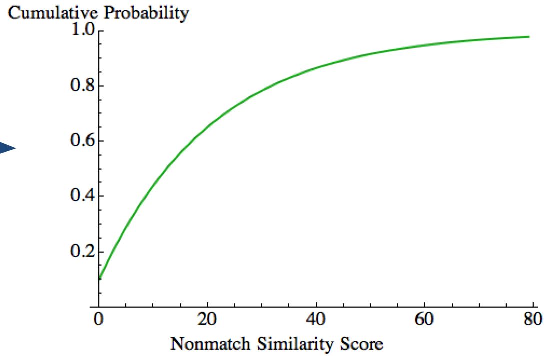
R(*s*): probability that an incorrect match will have similarity score less than *s*

Response Variable

$$q = \int_0^{\infty} Q(s)R(s)ds$$



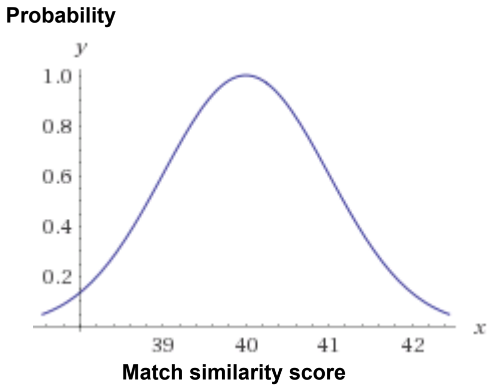




R(*s*): probability that an incorrect match will have similarity score less than *s*

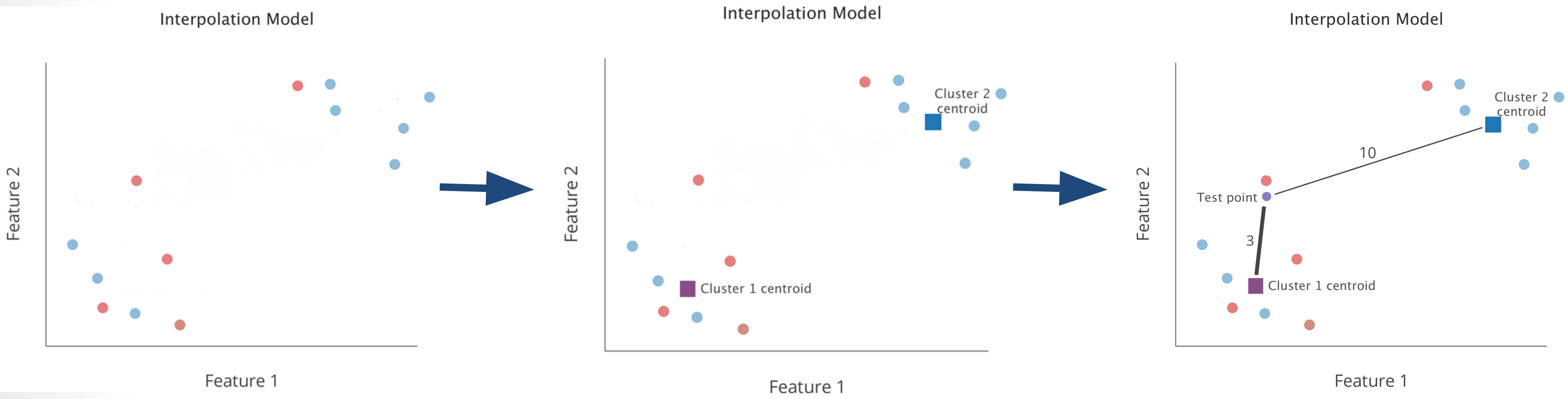
Response Variable

$$q = \int_0^\infty Q(s)R(s)ds$$

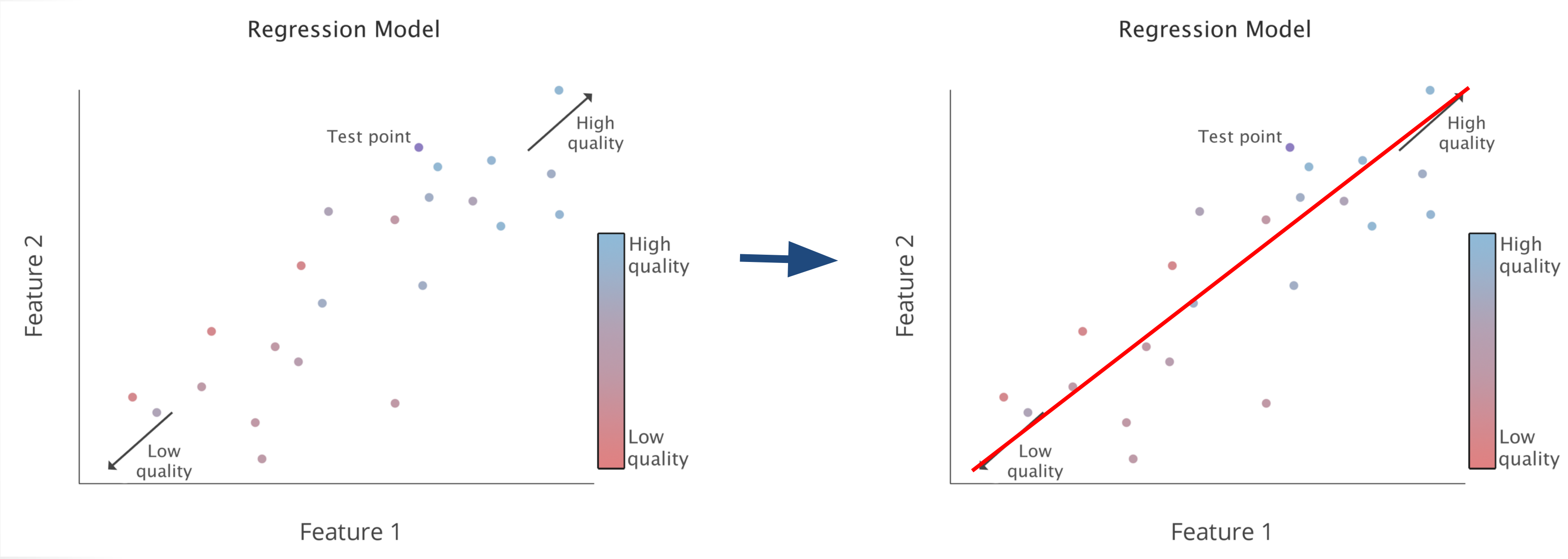


Q(*s*): probability that correct match will have similarity score *s*

Models: Clustering/Interpolation



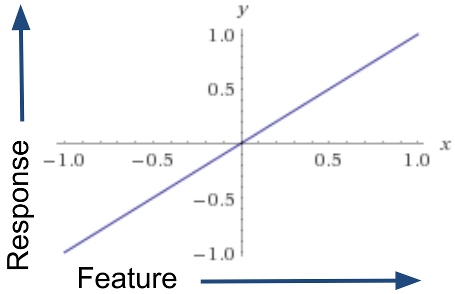
Models: Regression



Models

Linear Regression

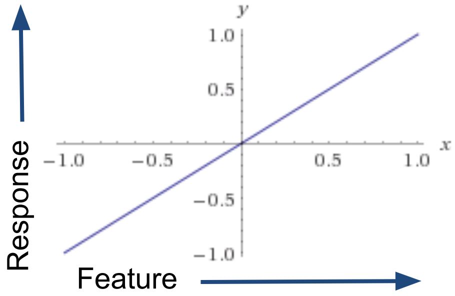
$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n$$



Models

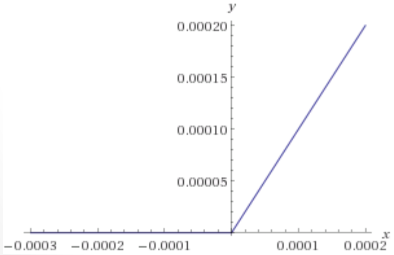
Linear Regression

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n$$



Capped Linear Regression

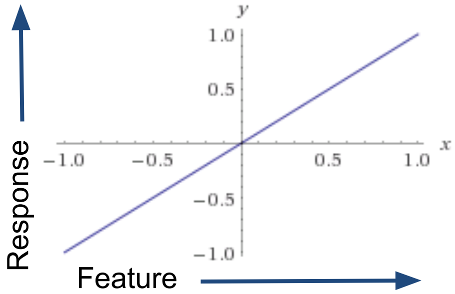
$$y = \max(b, a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n)$$



Models

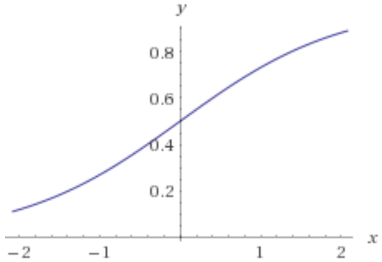
Linear Regression

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n$$



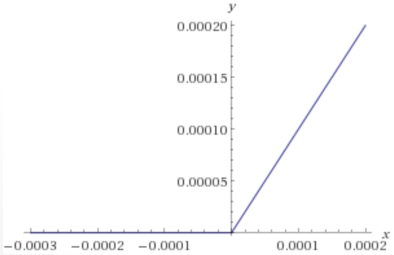
Logistic Regression

$$y = \frac{b}{1 + e^{a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n}}$$



Capped Linear Regression

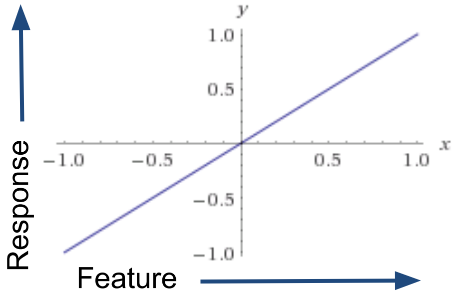
$$y = \max(b, a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n)$$



Models

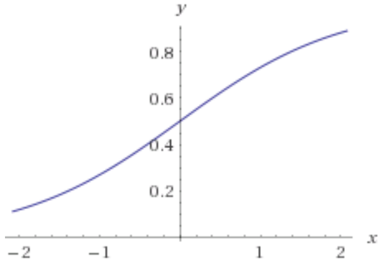
Linear Regression

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots a_nx_n$$



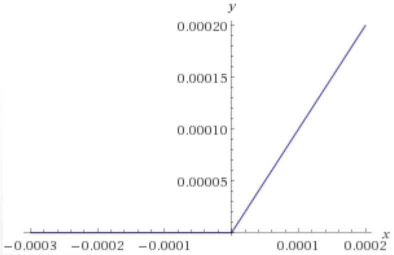
Logistic Regression

$$y = \frac{b}{1 + e^{a_0 + a_1x_1 + a_2x_2 + \ldots a_nx_n}}$$



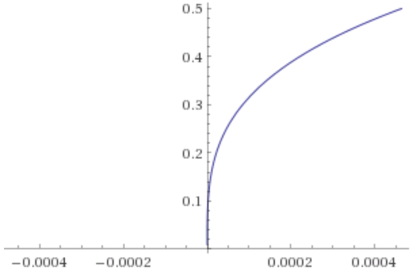
Capped Linear Regression

$$y = \max(b, a_0 + a_1x_1 + a_2x_2 + \ldots a_nx_n)$$

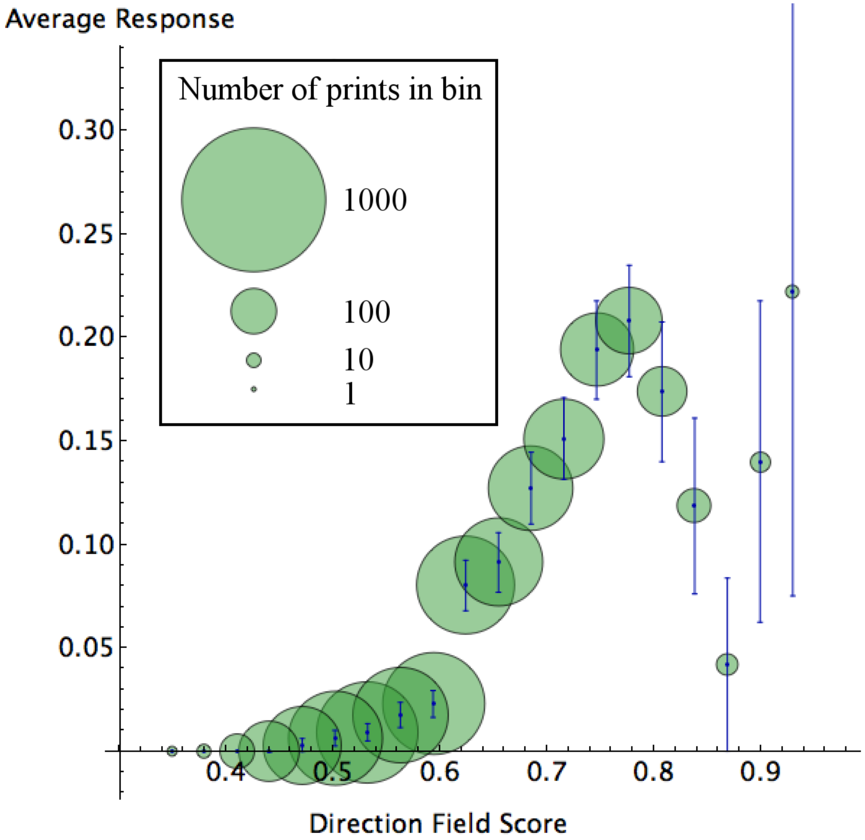


“Product” regression

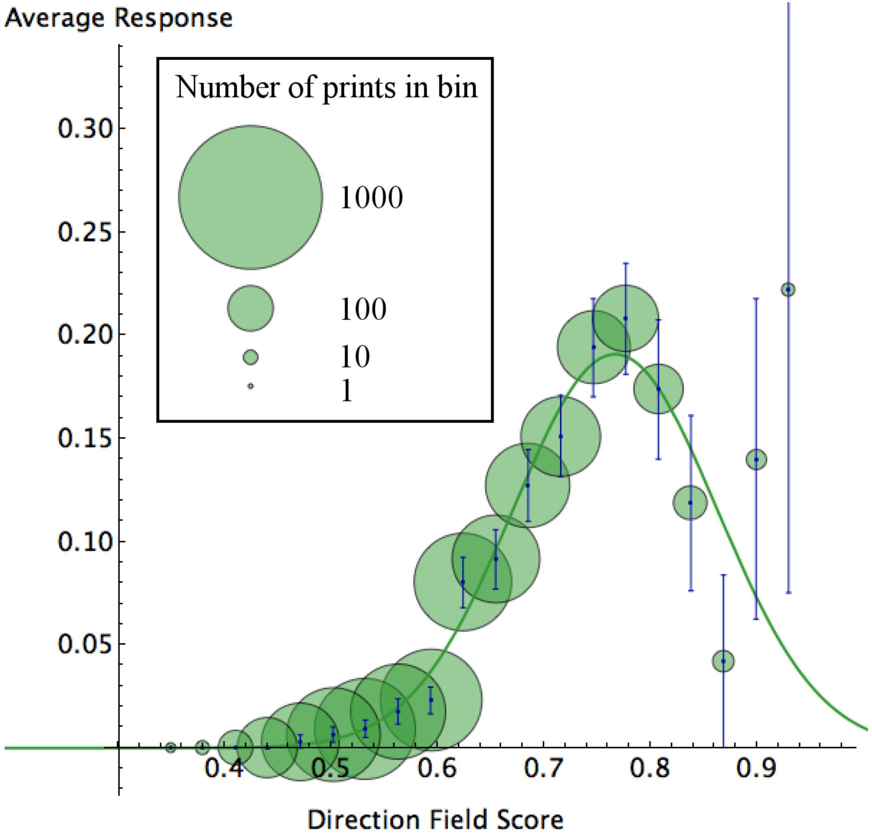
$$y = a(x_1x_2 \ldots x_n)^b$$



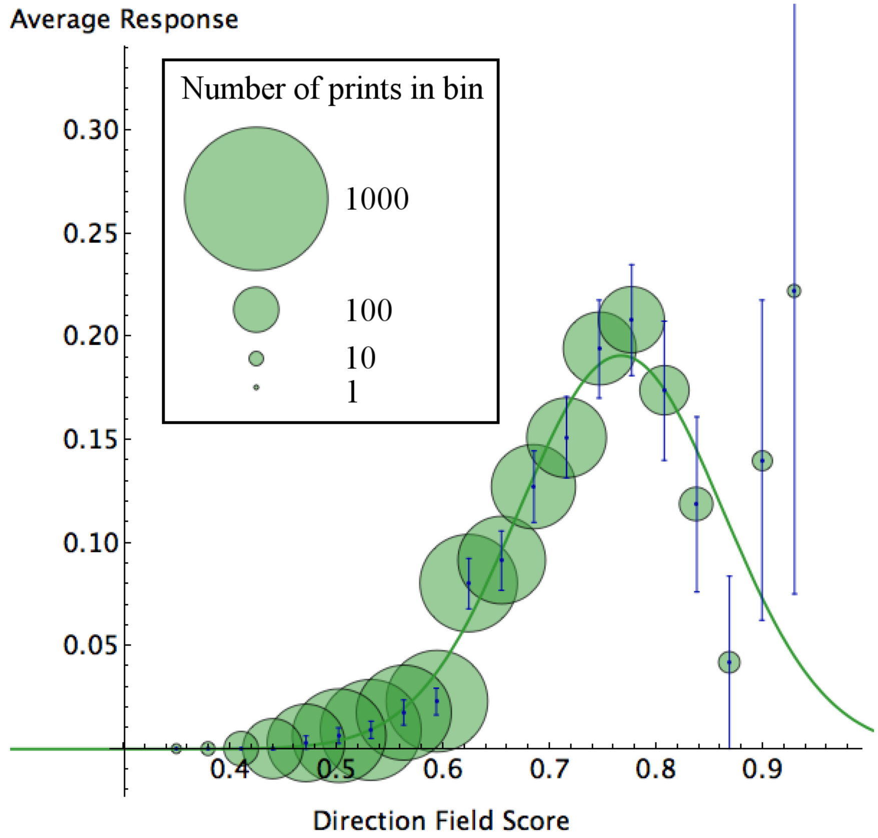
Feature Transformations

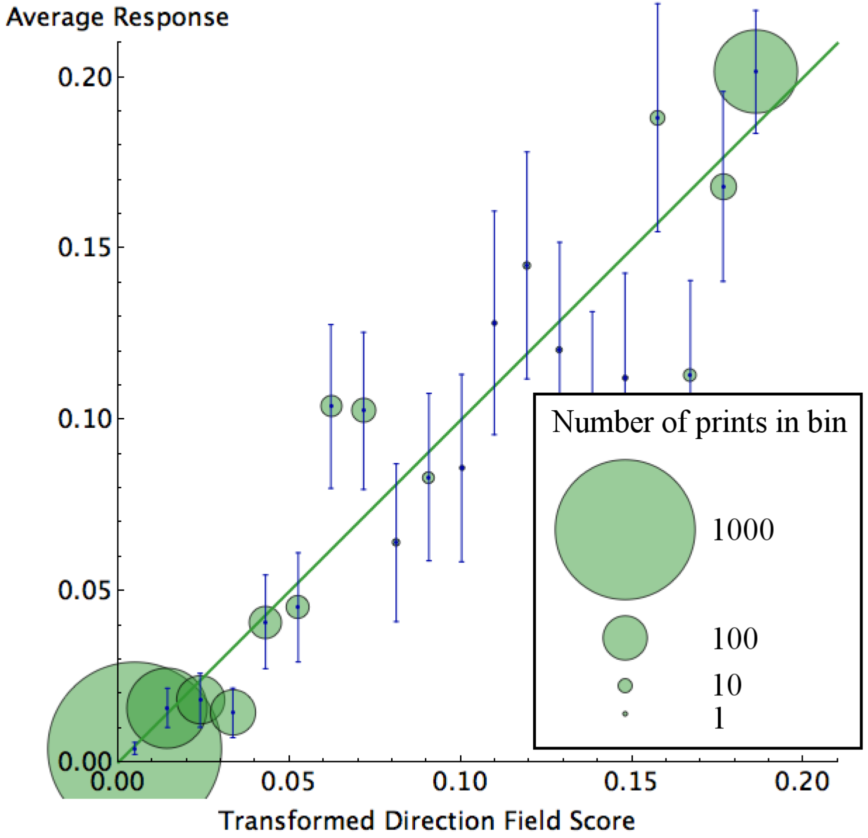


Feature Transformations

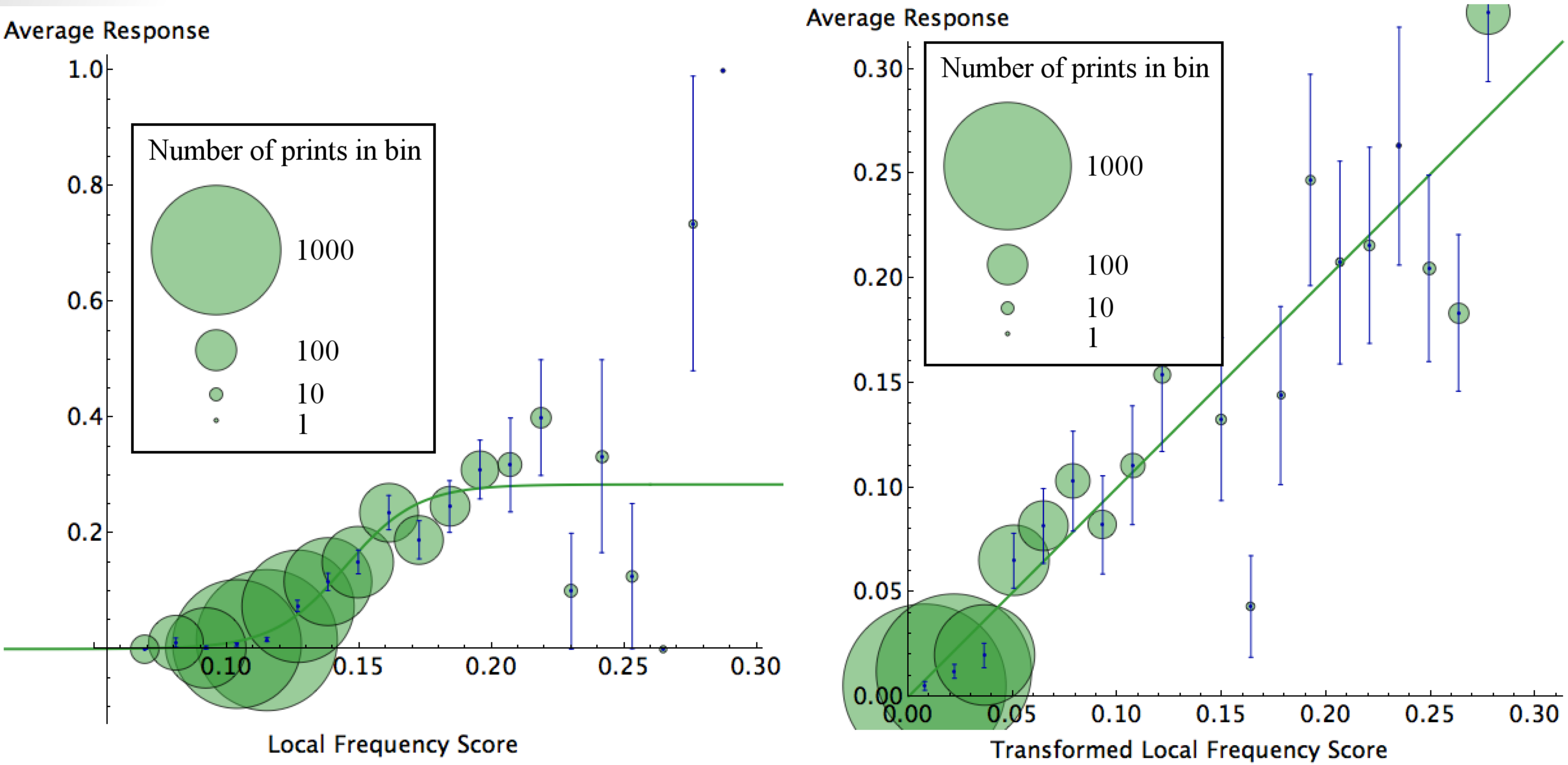


Feature Transformations





Feature Transformations



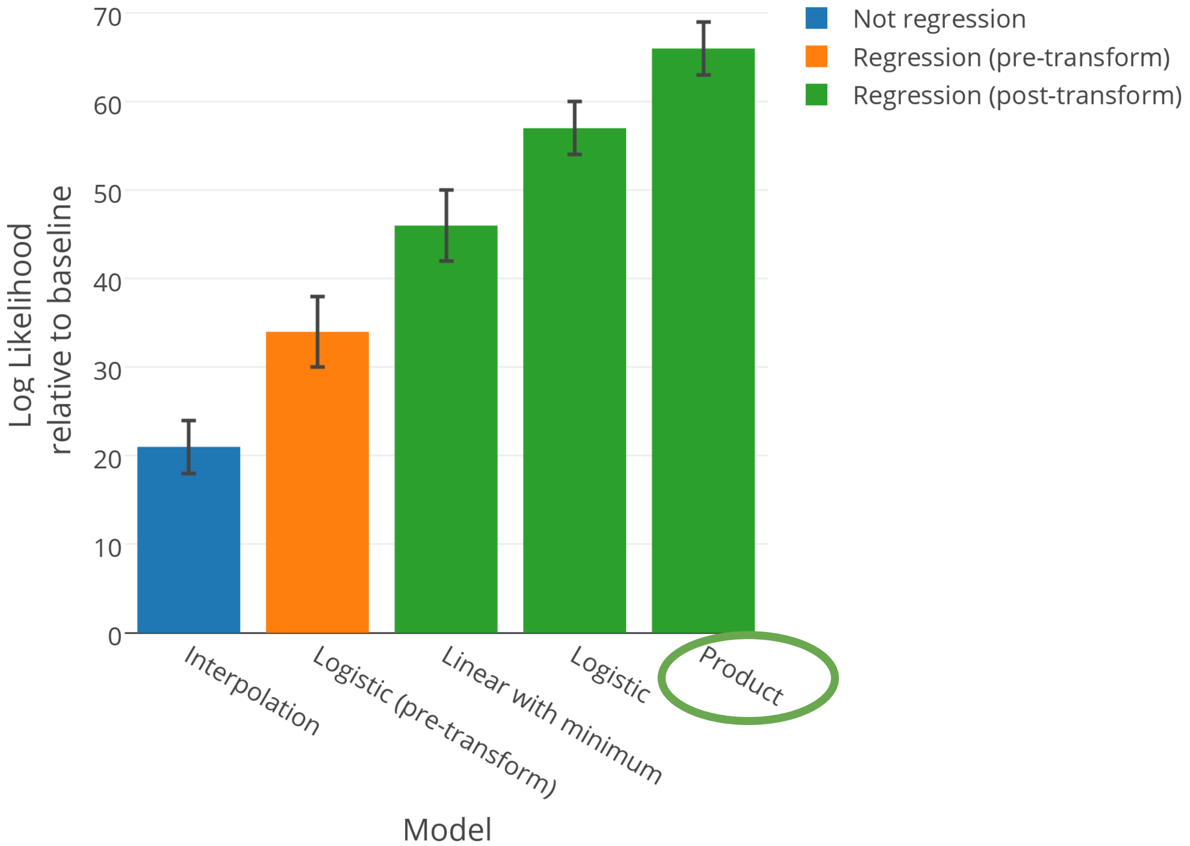
Log Likelihood

Our best statistic for judging a model:

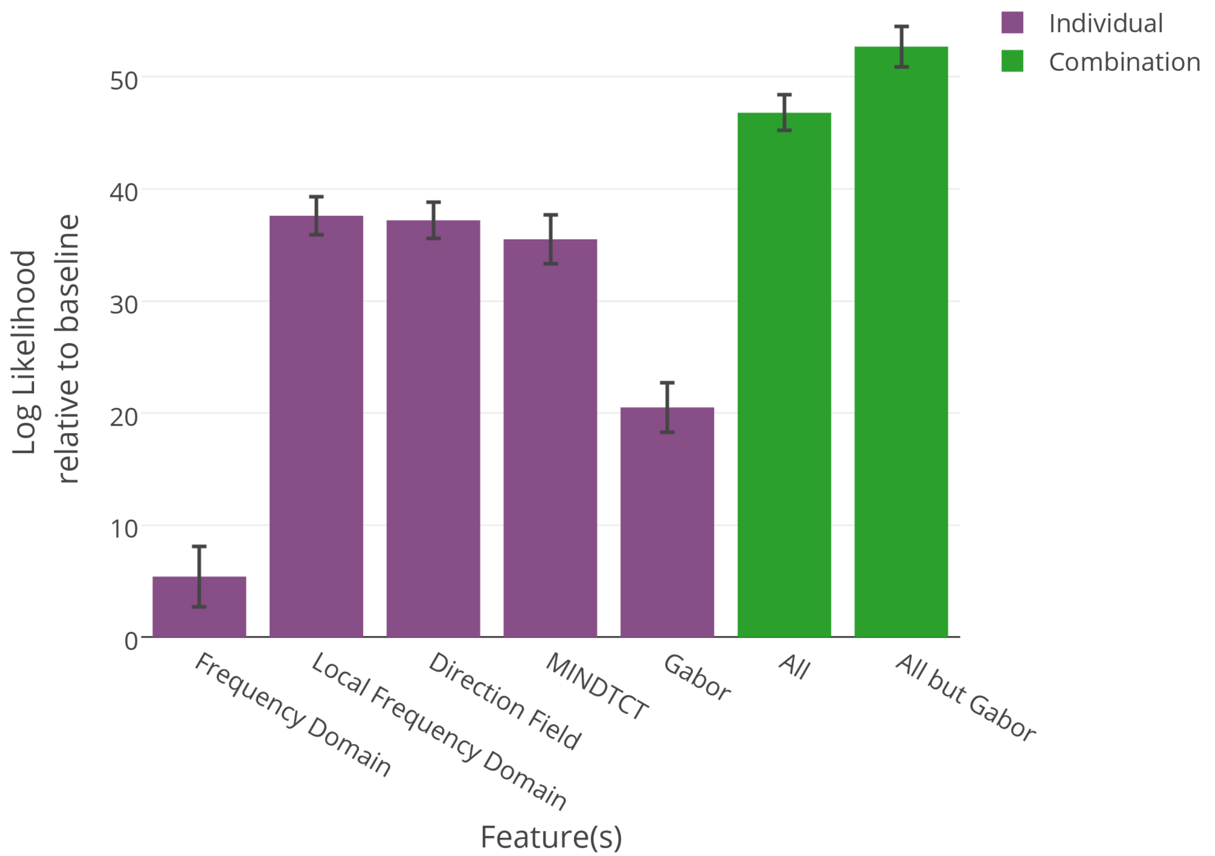
$$\text{log likelihood} = \sum_{\text{testing}} \ln(q_o q_p + (1 - q_0)(1 - q_p))$$

- The probability of observing testing data, assuming our model is correct
- Incorporates both how powerful the model is and how consistent its claims are
- The higher, the better

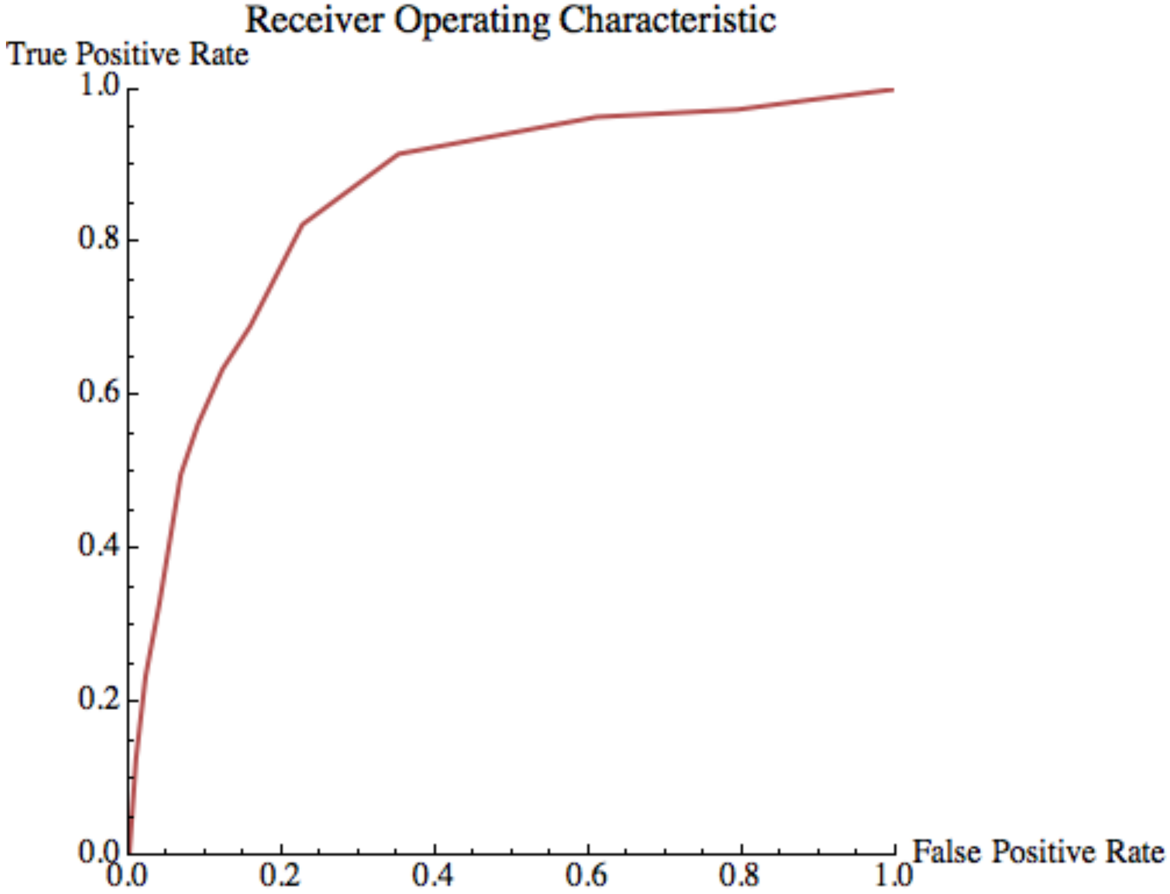
Model Performance



Feature Performance



High/Low Quality Classification



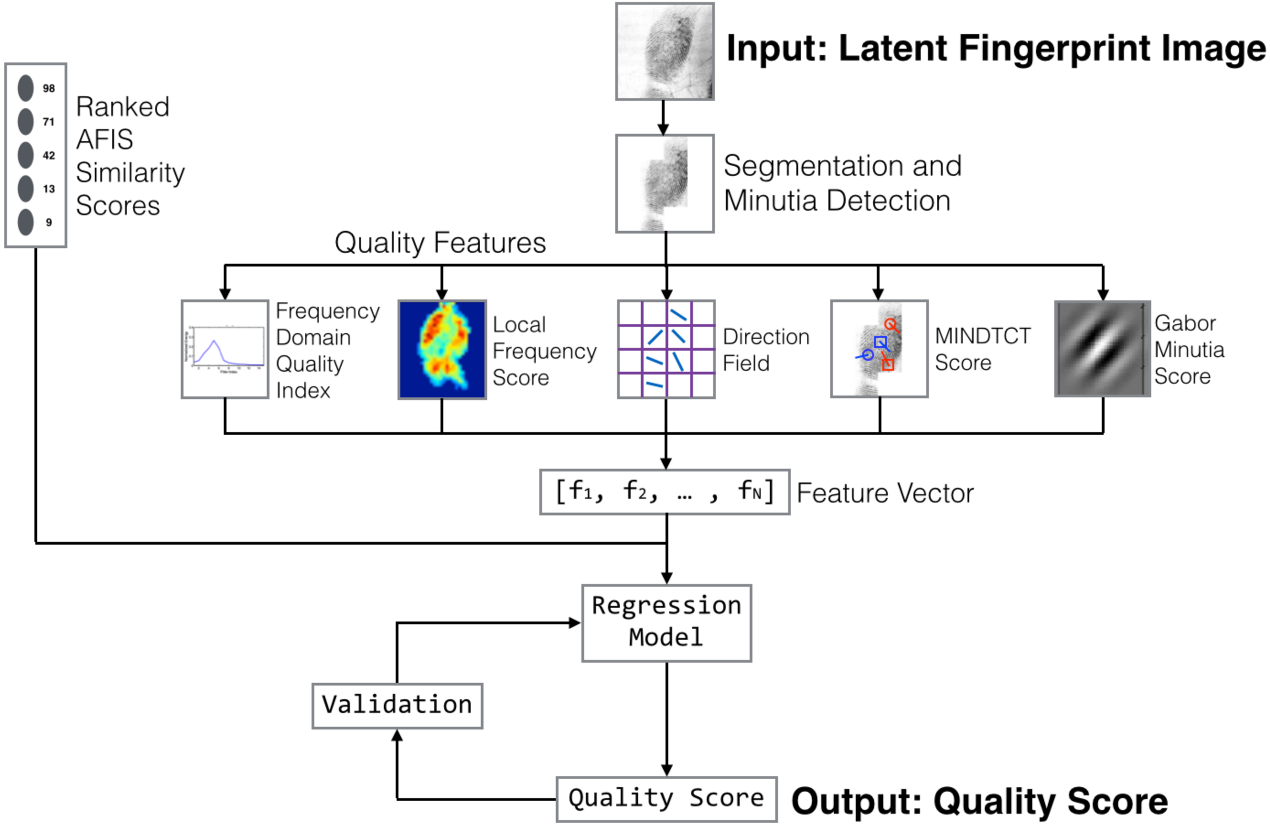
Limitations

- Only one data set of ~5000 latent prints and 120 exemplar prints used
- Data set prints only from 6 individuals
- Only one AFIS

Takeaways

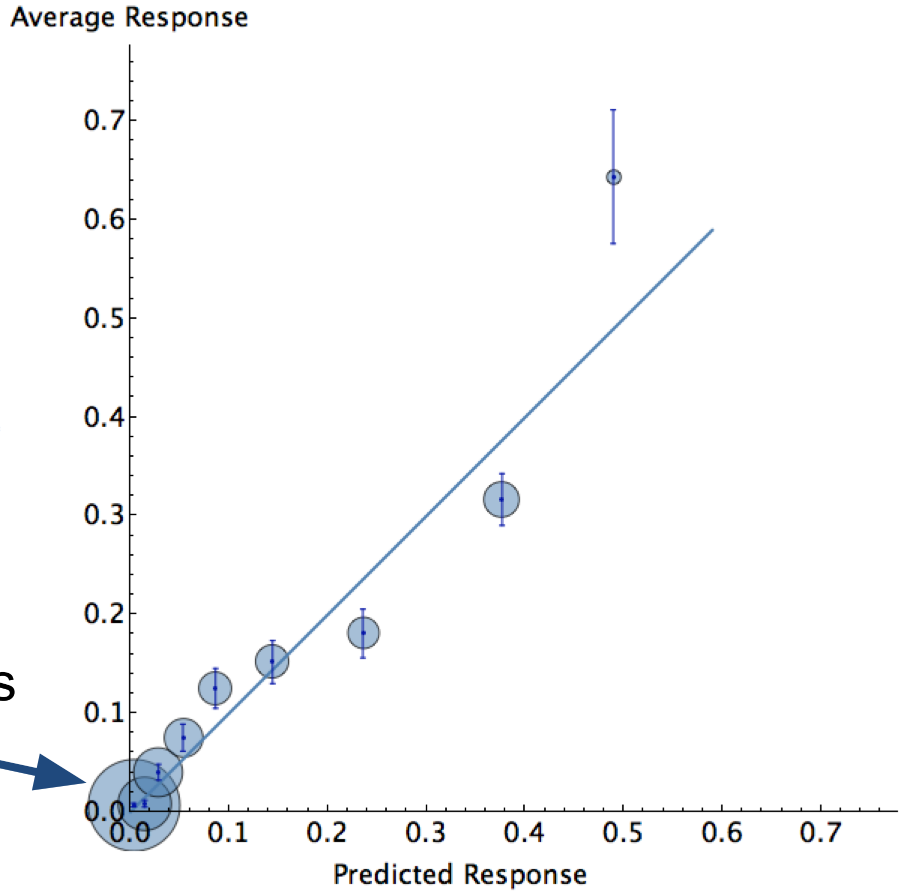
- Calculate quality of a print using a trained model
- Determined a model which effectively incorporates data from multiple features
- Reject at least 36% of latent prints with over 99% confidence

Questions?



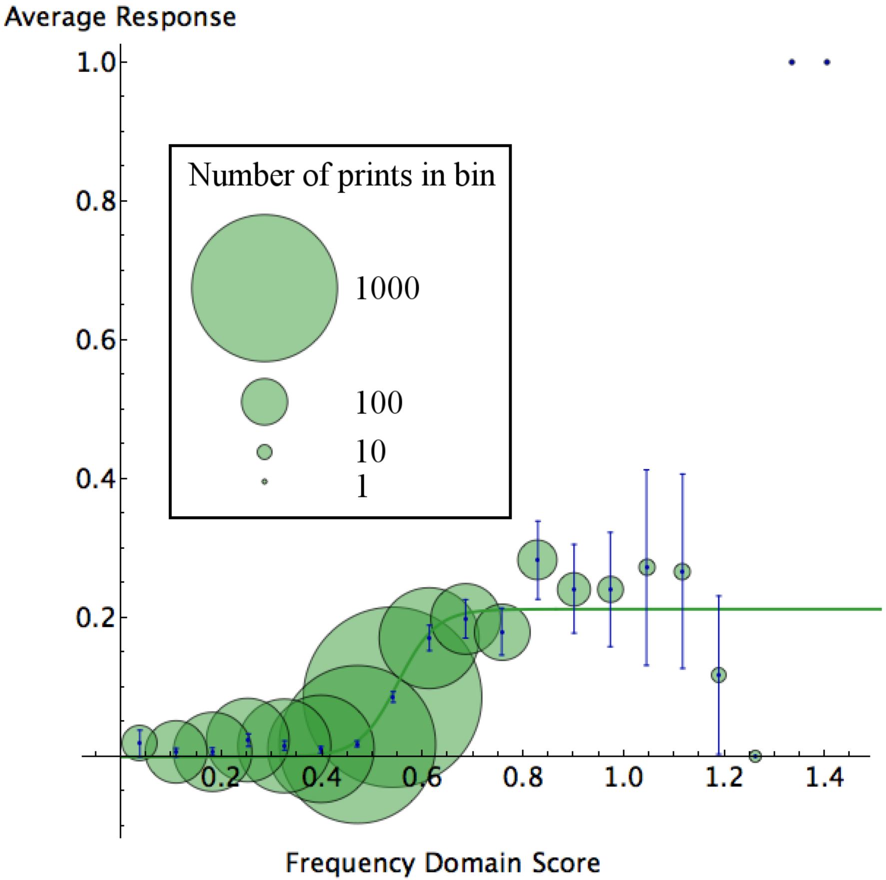
Input: Latent Fingerprint Image
Segmentation and Minutia Detection
Ranked AFIS Similarity Scores
98
71
42
13
9
Quality Features
Frequency Domain Quality Index
Local Frequency Score
Direction Field
MINDTCT Score
Gabor Minutia Score
[f1, f2, … , fN]
Feature Vector
Regression Model
Validation
Quality Score
Output: Quality Score

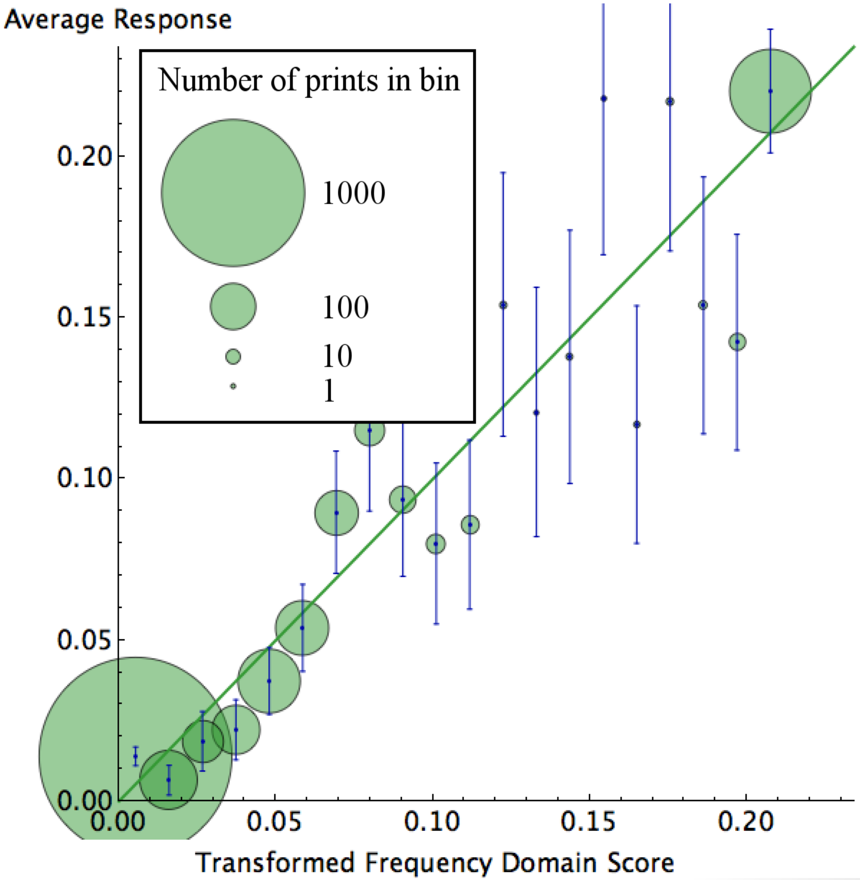
Model Predictions versus Response Variable



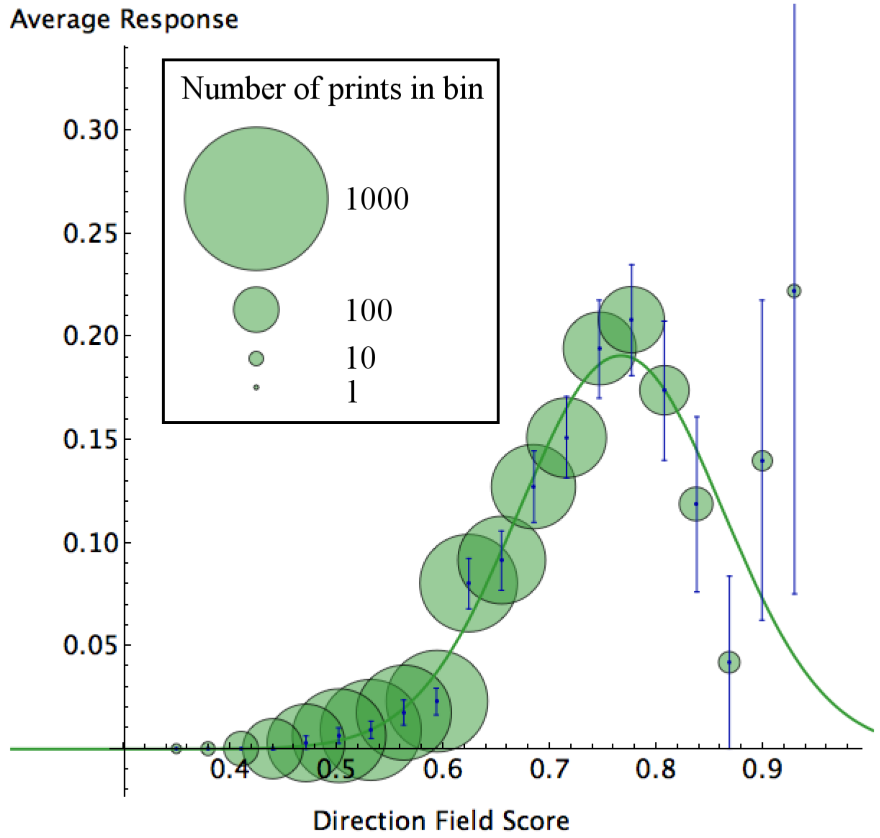
39% of test prints

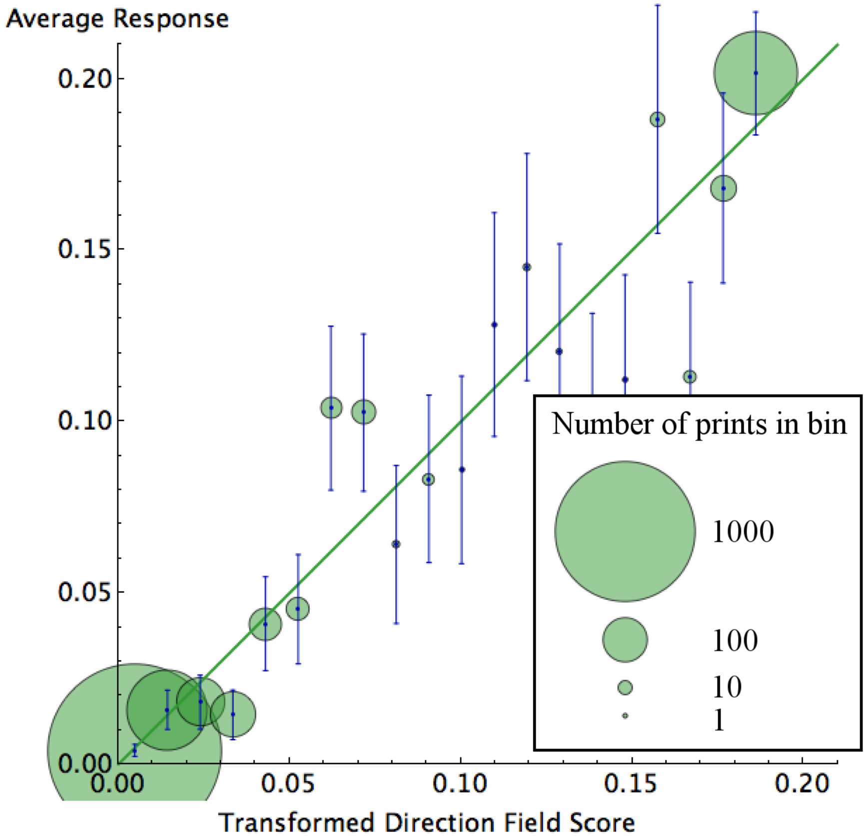
Feature Transformations



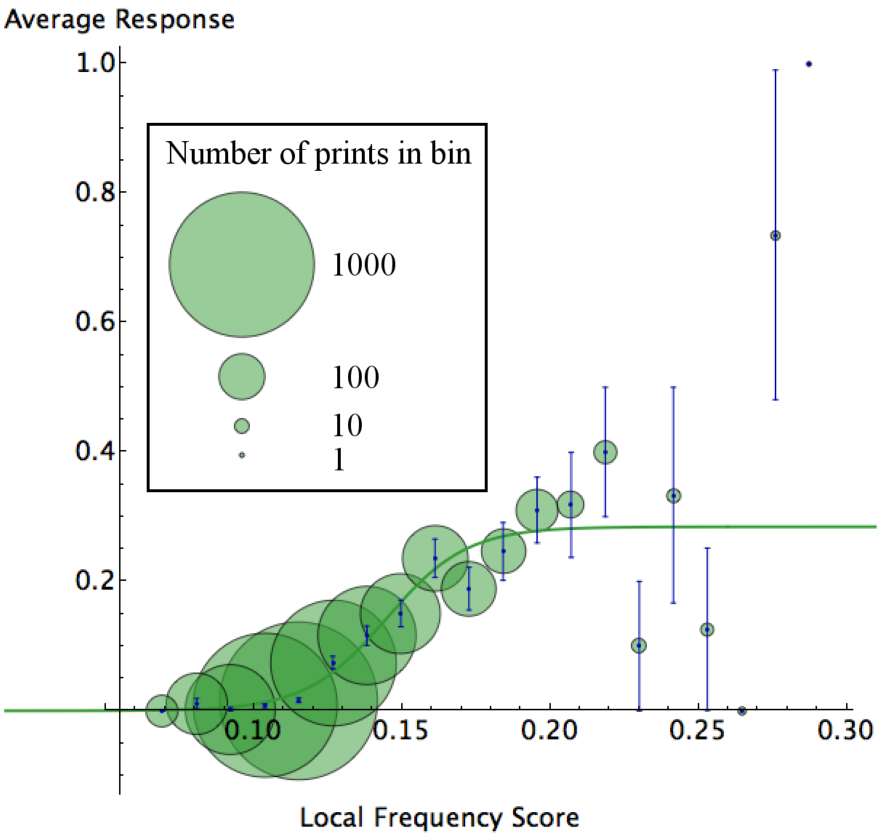


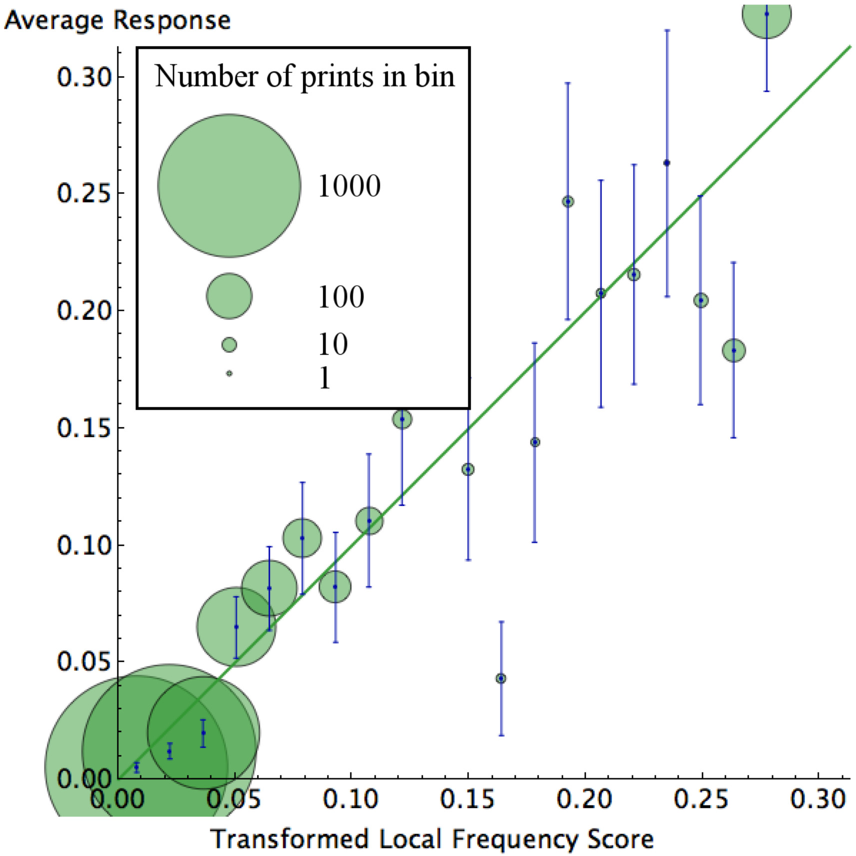
Feature Transformations



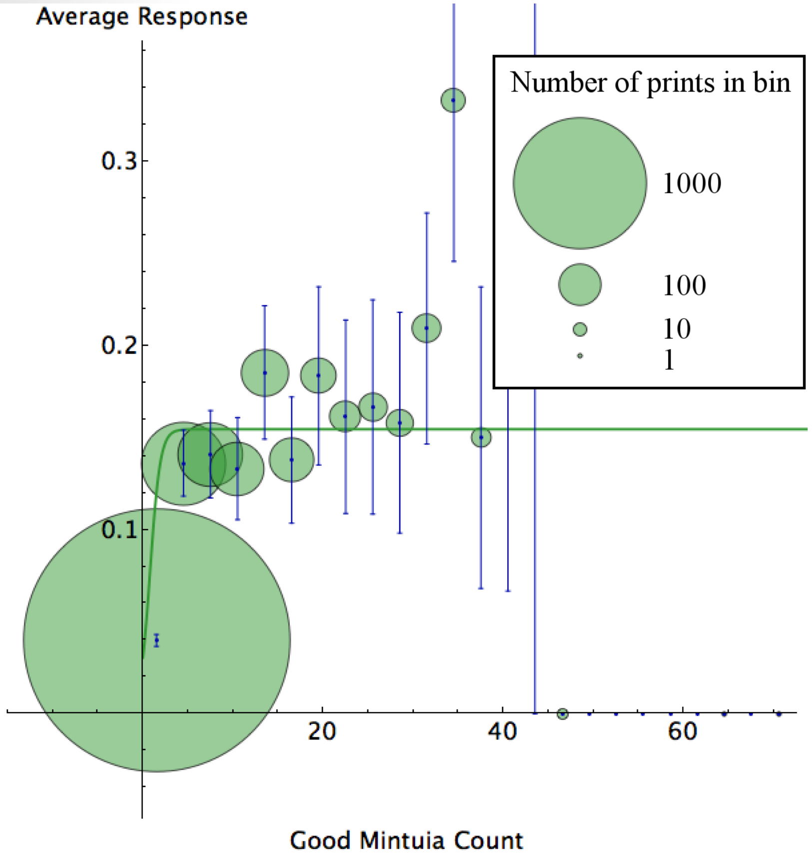


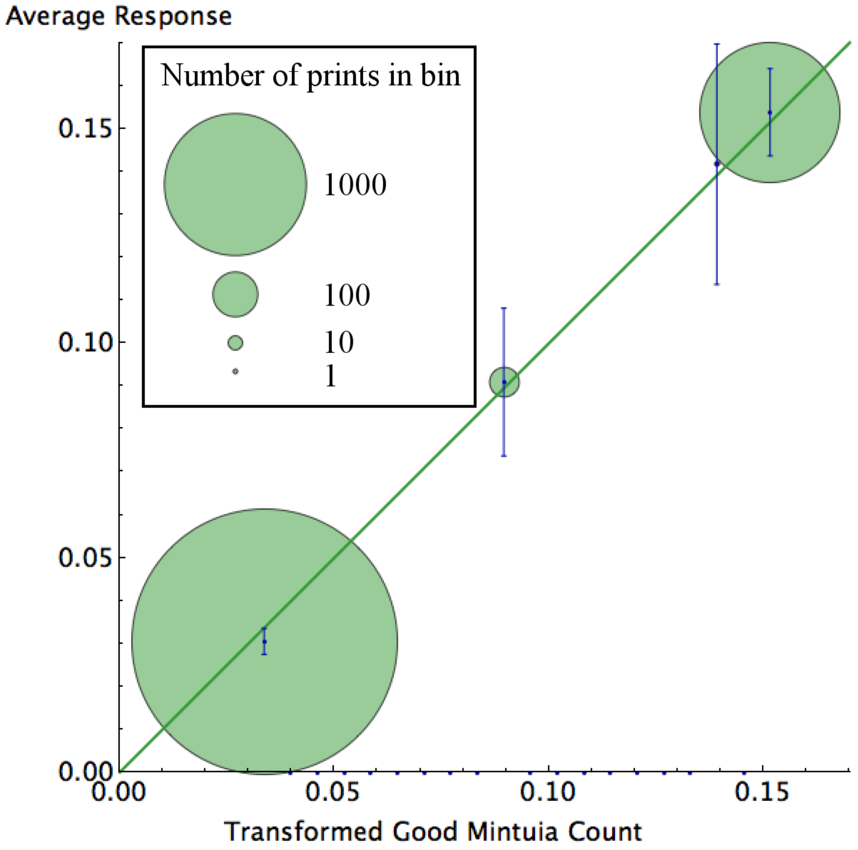
Feature Transformations



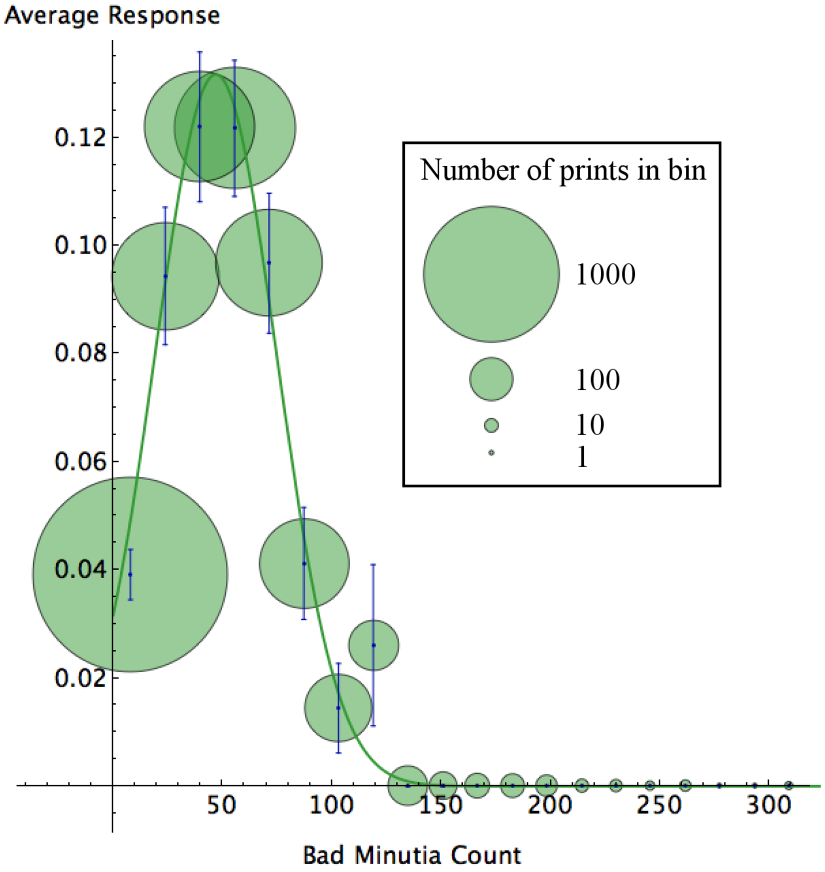


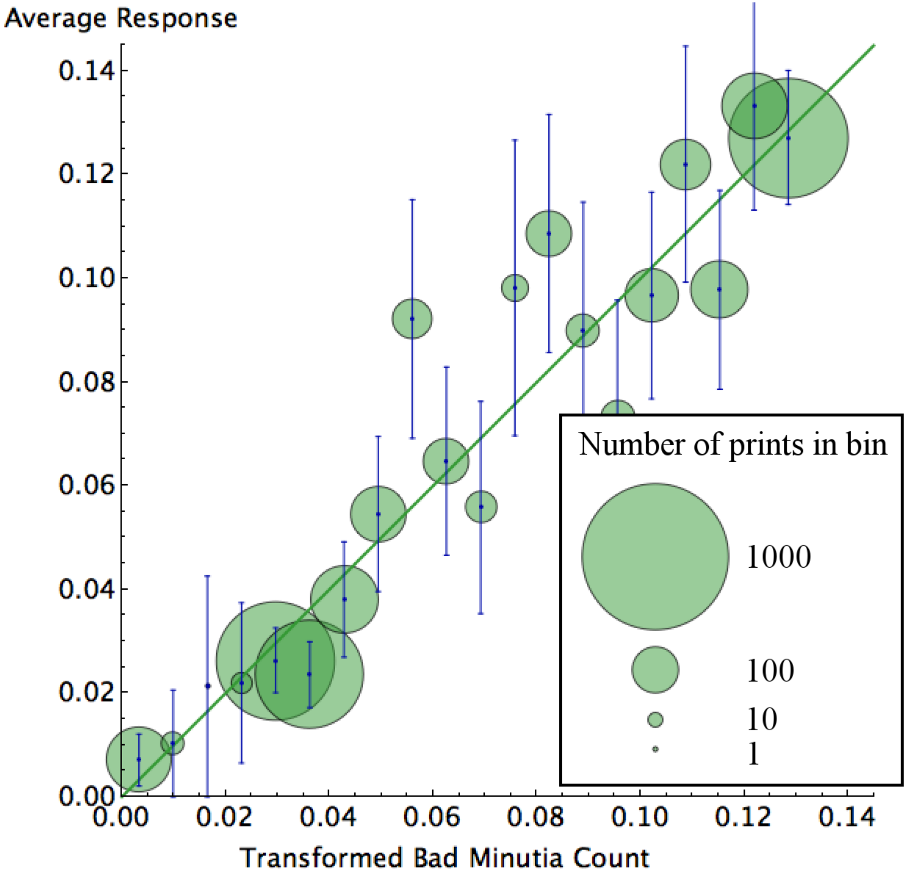
Feature Transformations



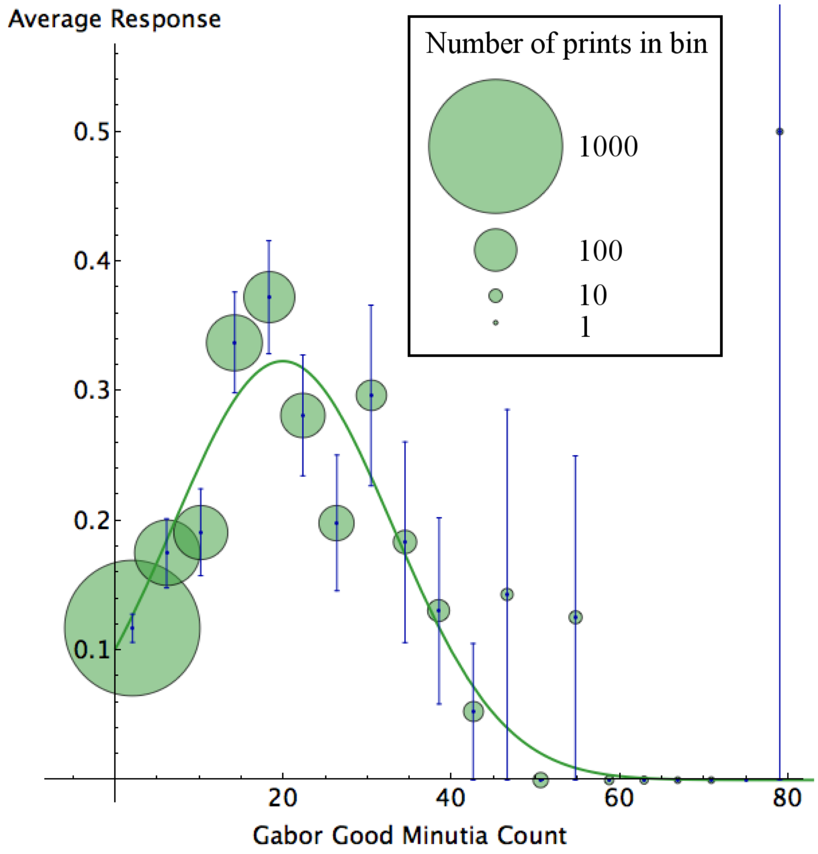


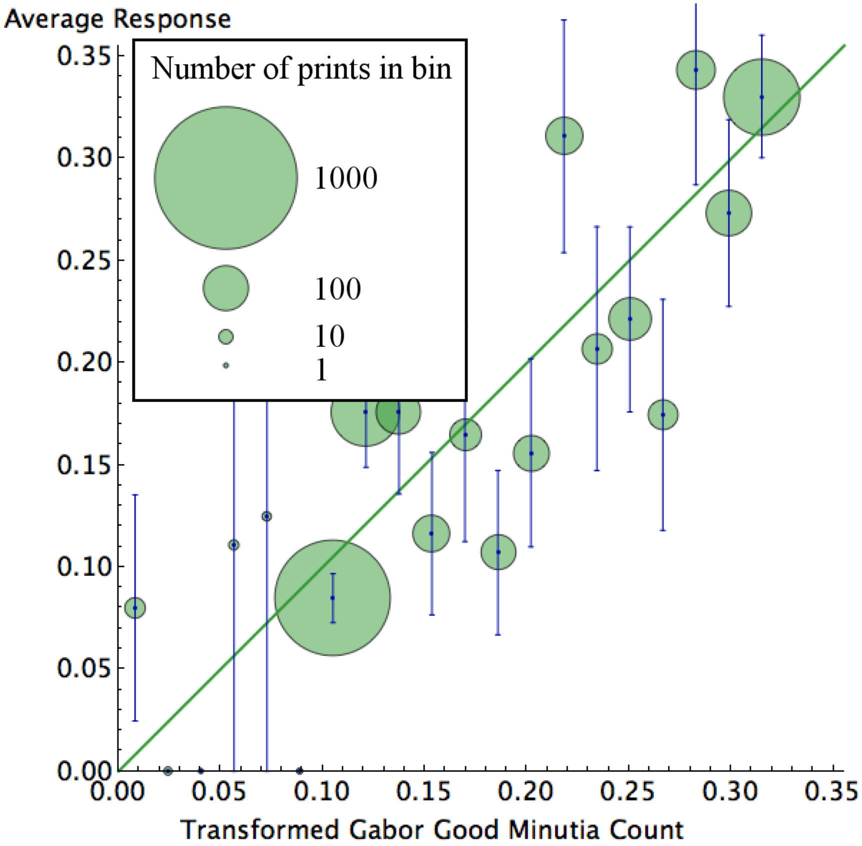
Feature Transformations





Feature Transformations





Feature Transformations