DePaul University
College of Computing and Digital Media

Casey Bennett, PhD

Feb. 4, 2019

Last Week

- Project Proposals

- In-Class Presentations
  - Online students, let me know by next Sunday Oct.27

PCA and Feature Selection

https://pollev.com/caseybennett801

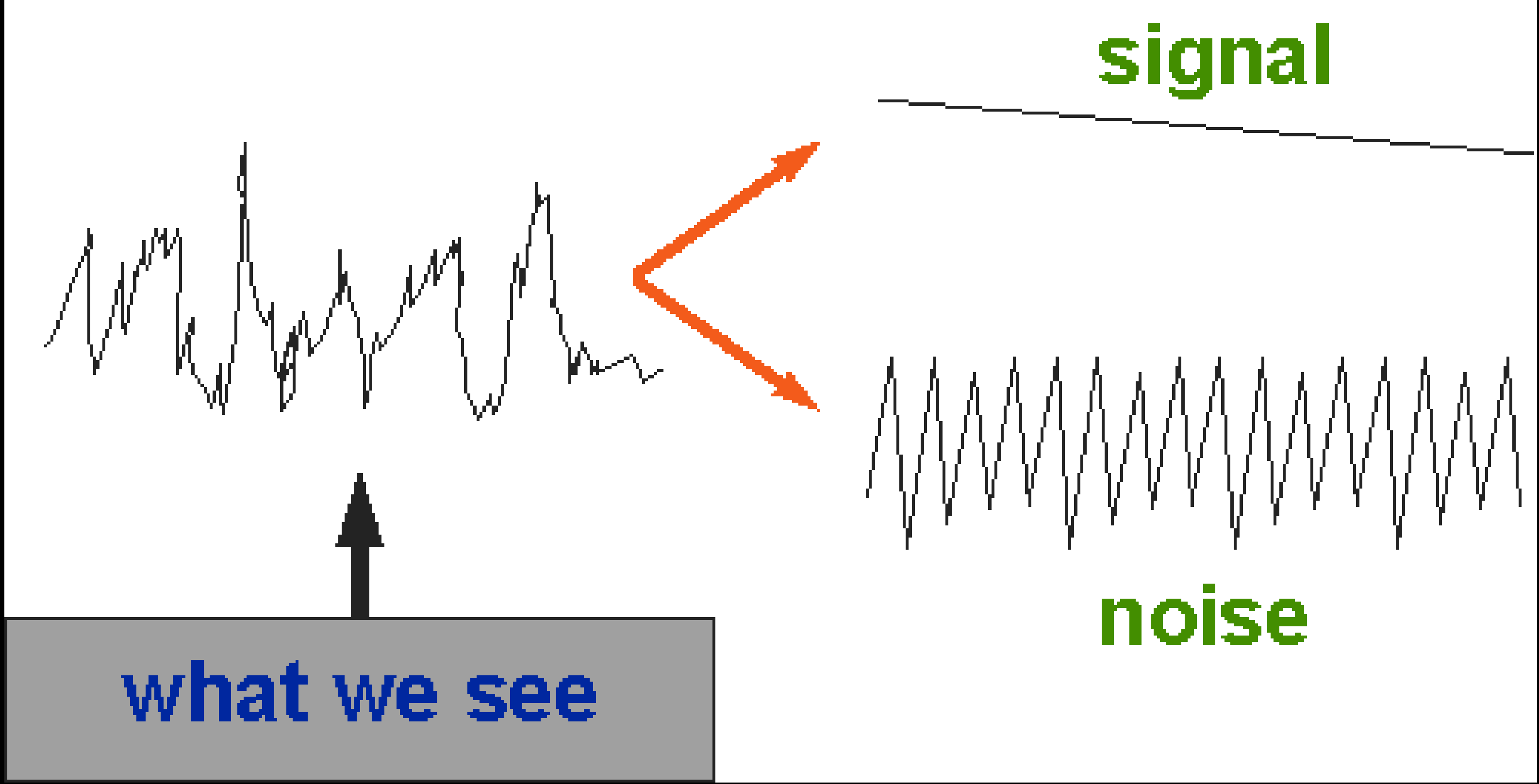
or text “caseybennett801” to 37607

“In the action, immediately look for the target, in words, listen closely to what is being signaled.”

– Marcus Aurelius

What does passage mean? Why did I put it up here?

What we observe can be divided into:



FACS - Ekman

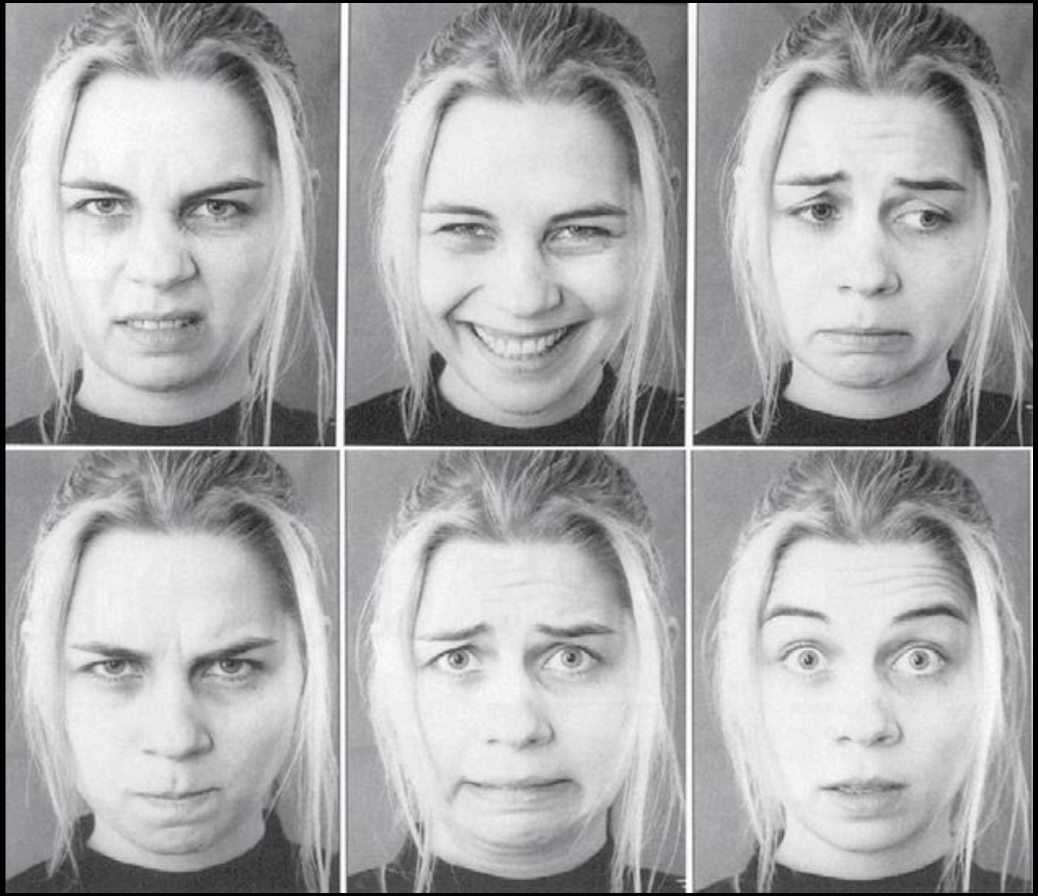
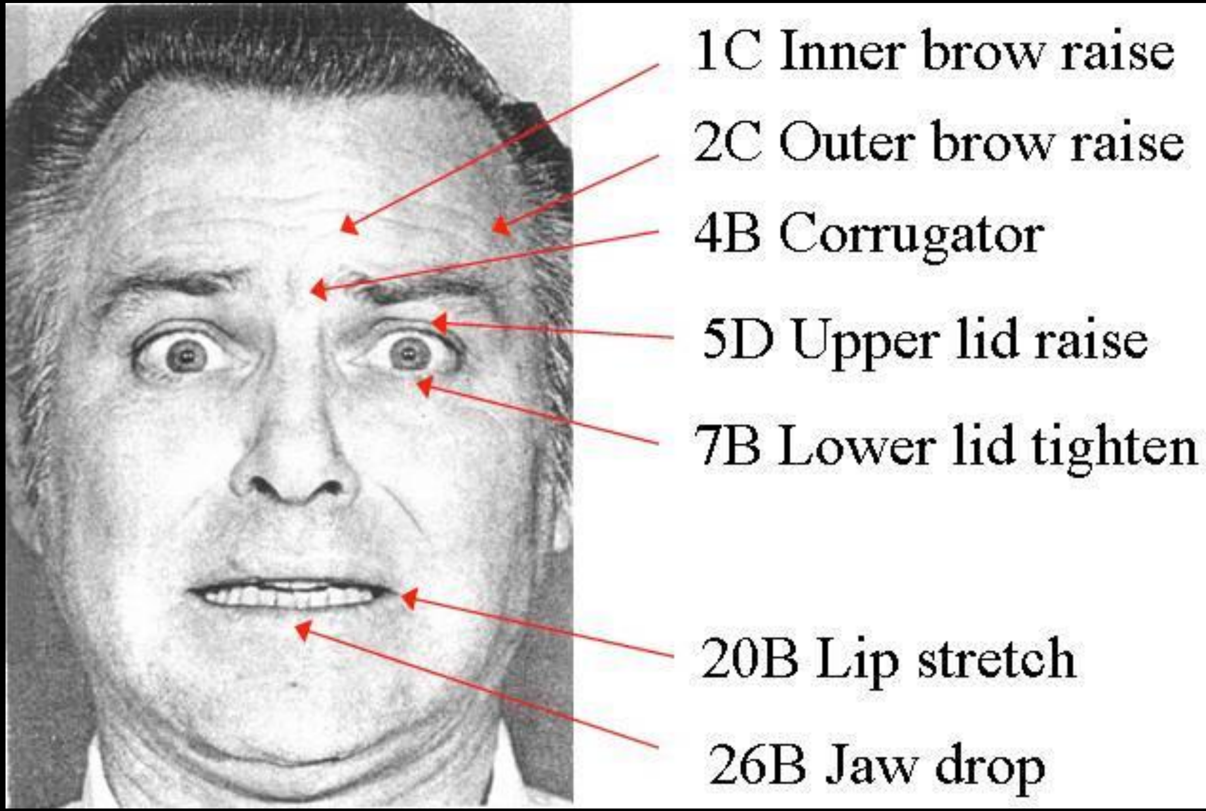


Figure 3. Prototypic facial expressions of six basic emotions (a–f): disgust, happiness, sadness, anger, fear and surprise.

From: Pantic (2009) Machine Analysis of Facial Behavior: Naturalistic and Dynamic Behavior



From: Littlewort et al. (2007) Faces of Pain: Automated Measurement of Spontaneous Facial Expressions of Genuine and Posed Pain

Feature Selection

10X Cross-Val (partitioned)					
Model	Binning	# of Features	Accuracy	Correct	Incorrect
Naïve Bayes	CAIM	5	87.7%	157	22
Naïve Bayes	CAIM	8	95.0%	170	9
Naïve Bayes	CAIM	10	96.7%	173	6
Naïve Bayes	CAIM	15	97.8%	175	4
Naïve Bayes	CAIM	All	97.2%	174	5
Naïve Bayes	None	5	86.6%	155	24
Naïve Bayes	None	8	98.3%	176	3
Naïve Bayes	None	10	98.9%	177	2
Naïve Bayes	None	15	98.3%	176	3
Naïve Bayes	None	All	97.8%	175	4
Bayes Net - K2	CAIM	8	95.0%	170	9
Bayes Net - K2	CAIM	10	96.7%	173	6
Bayes Net - K2	CAIM	All	98.3%	176	3
Bayes Net - K2	None	8	92.2%	165	14
Bayes Net - K2	None	10	94.4%	169	10
Bayes Net - K2	None	All	96.7%	173	6
Ensemble	CAIM	8	93.9%	168	11
Ensemble	CAIM	10	96.1%	172	7
Ensemble	CAIM	All	97.8%	175	4
Ensemble	None	8	96.1%	172	7
Ensemble	None	10	96.6%	173	6
Ensemble	None	All	95.5%	171	8

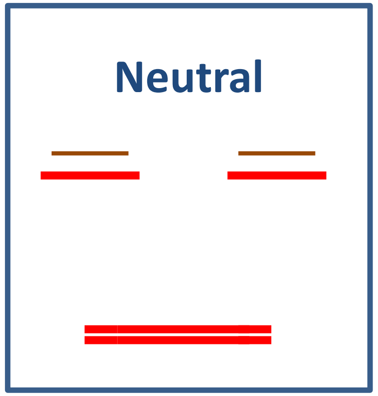
Bayes Net- K2	None	Chi-Squared
Ranked attributes:		
246.573	AU6CheekRaise	CHEEK
204.387	AU12LipCornerPull	MOUTH
200.844	AU9NoseWrinkle	NOSE
195.337	AU15LipCornerDepresso	MOUTH
190.471	AU25LipsPart	MOUTH
179.358	AU24LipPresser	MOUTH
162.000	AU20LipStretch	MOUTH
147.320	AU17ChinRaise	CHIN
144.791	AU4BrowLower	EYE
139.803	AU1InnerBrowRaise	EYE

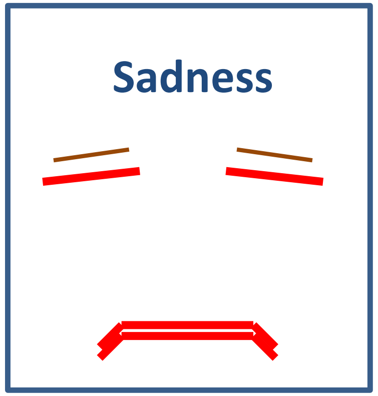
Bayes Net- K2	CAIM	Chi-Squared
Ranked attributes:		
262.867	AU12LipCornerPull	MOUTH
253.565	AU9NoseWrinkle	NOSE
252.381	AU6CheekRaise	CHEEK
230.397	AU25LipsPart	MOUTH
220.522	AU15LipCornerDepressc	MOUTH
209.801	AU20LipStretch	MOUTH
209.569	AU24LipPresser	MOUTH
198.623	AU1InnerBrowRaise	EYE
192.864	AU5EyeWiden	EYE
191.497	AU2OuterBrowRaise	EYE

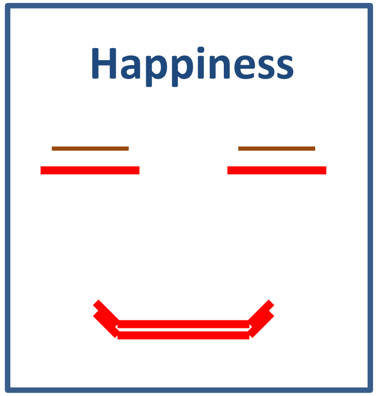
Ensemble	None	Chi-Squared
Ranked attributes:		
203.333	AU9NoseWrinkle	NOSE
202.730	AU12LipCornerPull	MOUTH
201.626	AU6CheekRaise	CHEEK
197.173	AU25LipsPart	MOUTH
178.658	AU24LipPresser	MOUTH
174.075	AU15LipCornerDepresso	MOUTH
162.000	AU20LipStretch	MOUTH
150.034	AU4BrowLower	EYE
143.030	AU17ChinRaise	CHIN
139.962	AU1InnerBrowRaise	EYE

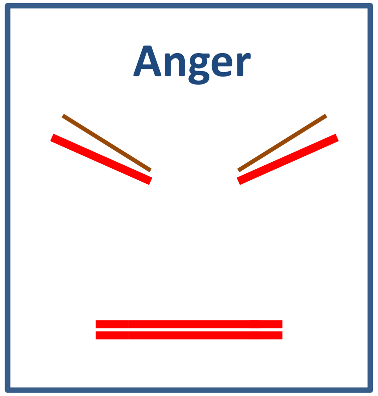
Ensemble	CAIM	Chi-Squared
Ranked attributes:		
276.639	AU12LipCornerPull	MOUTH
259.229	AU6CheekRaise	CHEEK
247.147	AU9NoseWrinkle	NOSE
216.852	AU25LipsPart	MOUTH
216.336	AU15LipCornerDepressc	MOUTH
211.459	AU20LipStretch	MOUTH
203.897	AU4BrowLower	EYE
203.433	AU24LipPresser	MOUTH
199.712	AU5EyeWiden	EYE
189.544	AU2OuterBrowRaise	EYE

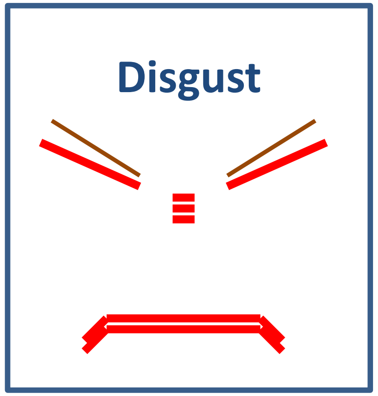
Facial Schematics

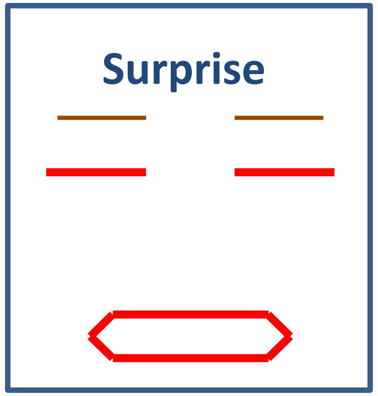


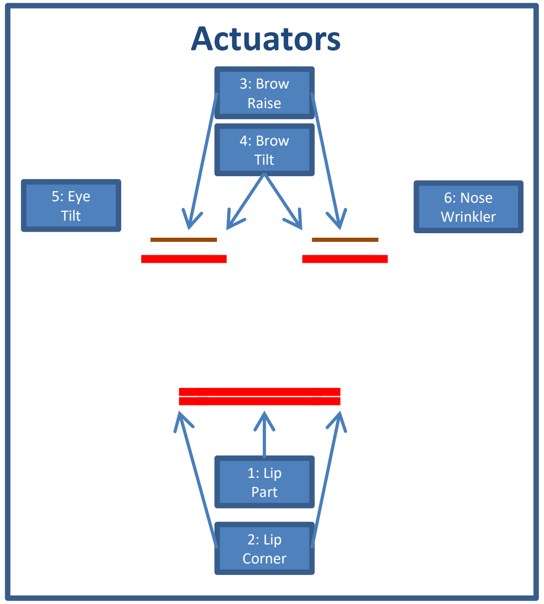




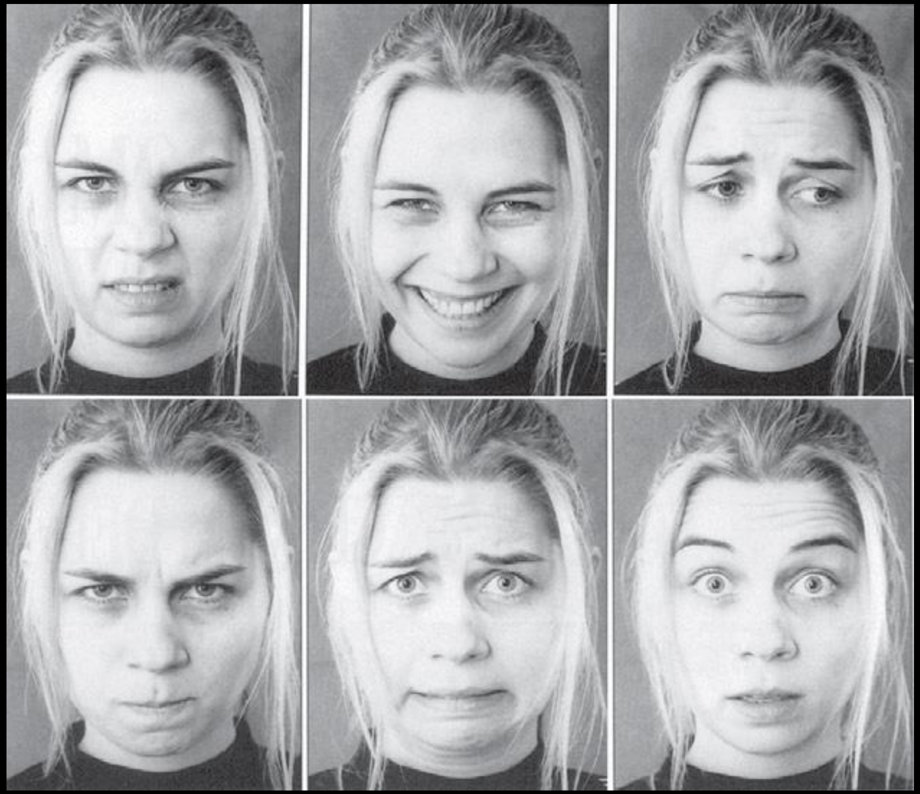




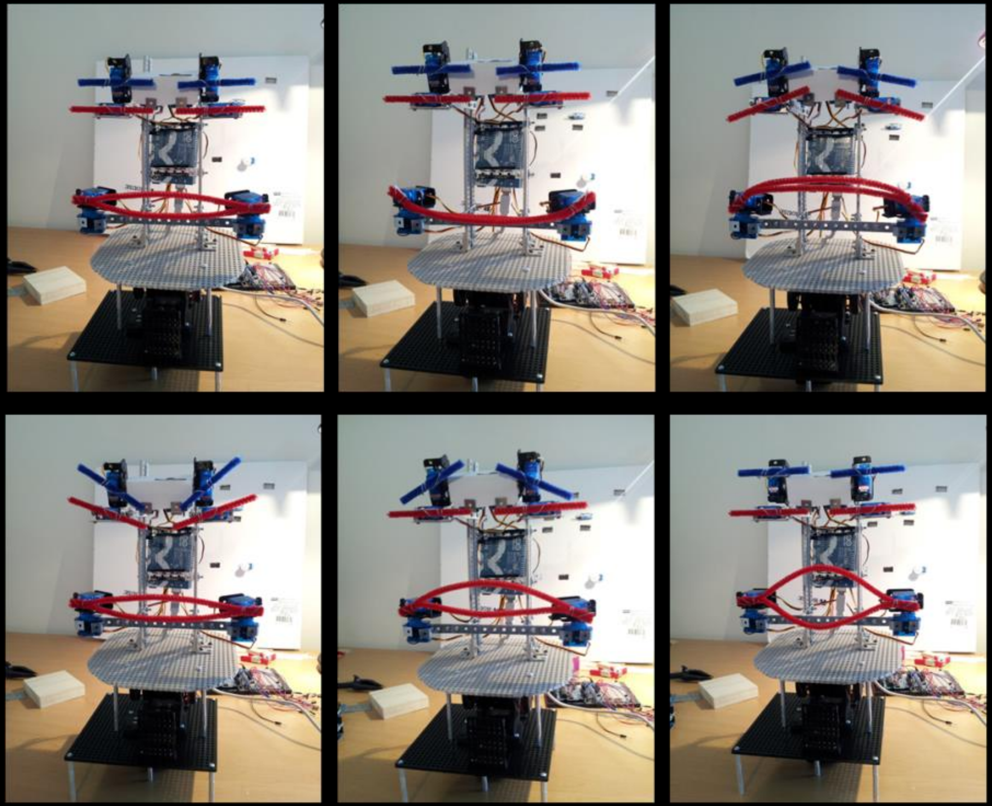


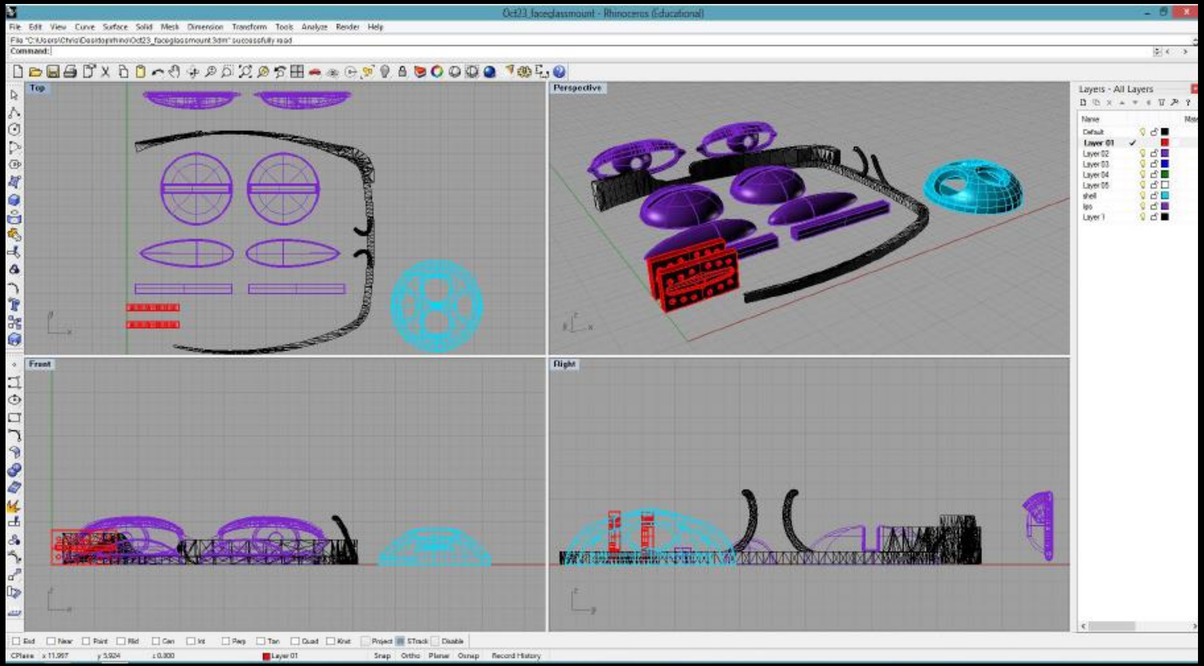


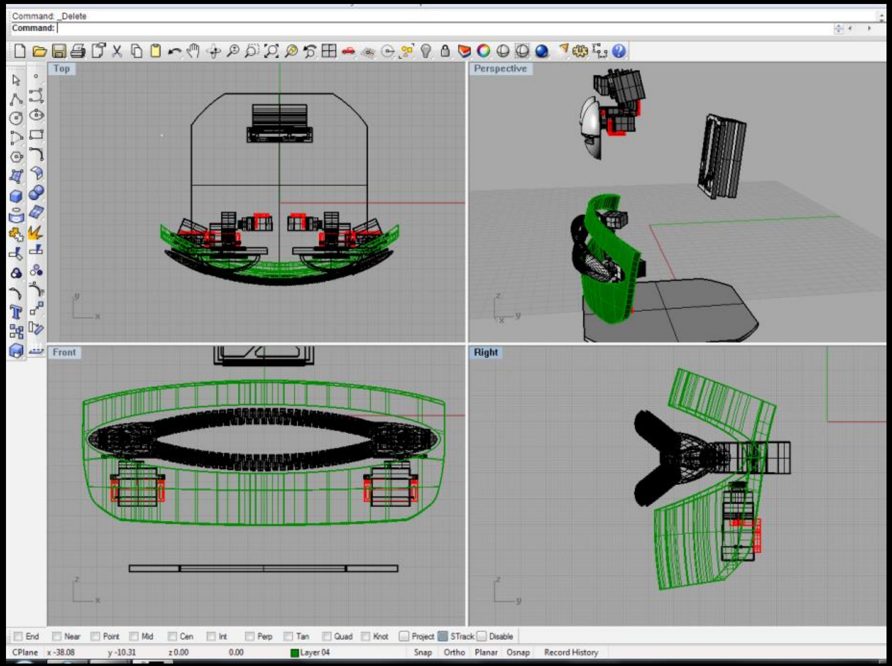
Start with the Minimalist Design

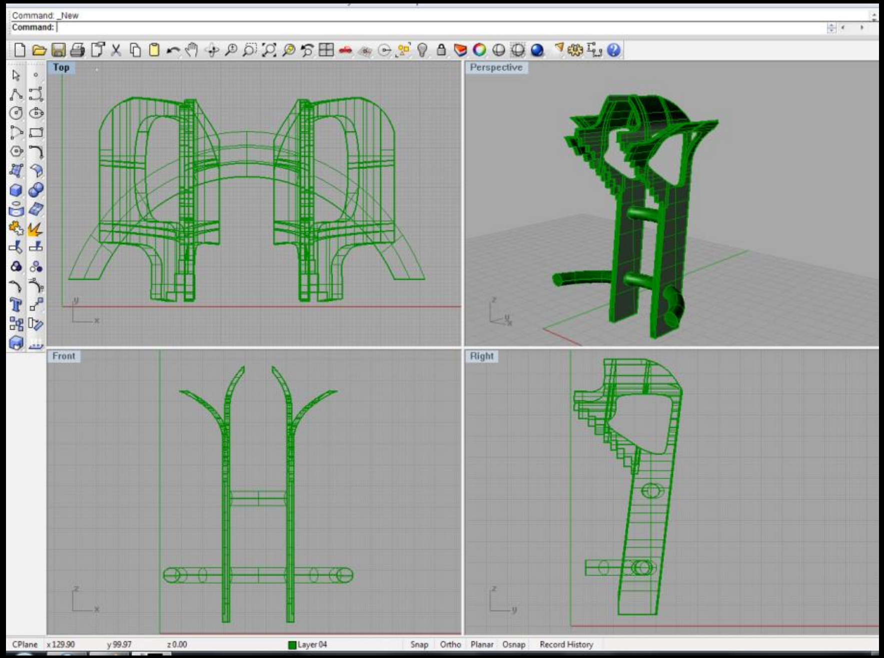


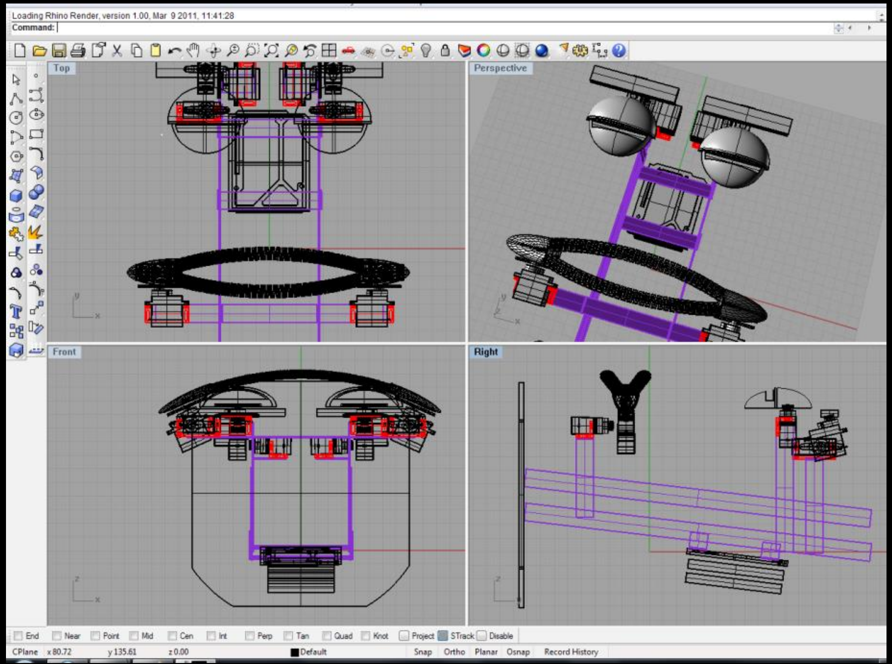
From: Pantic (2009) “Machine analysis of Facial behavior: Naturalistic and dynamic behavior. “



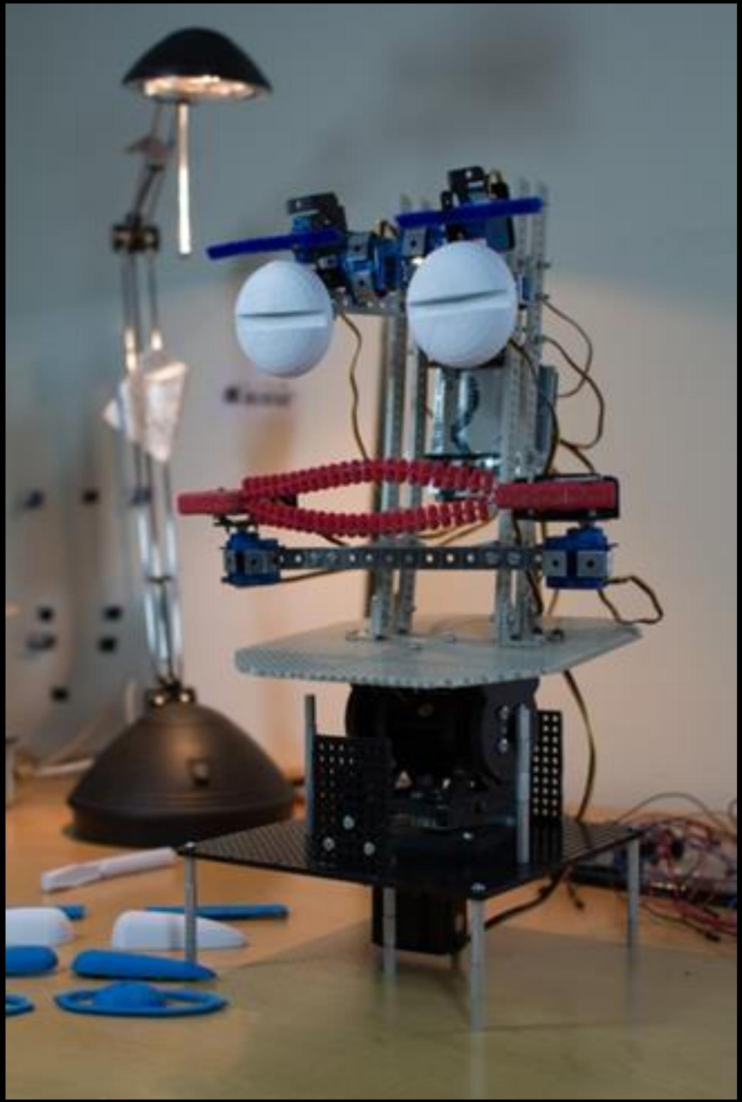

Oct23_faceglassmount - Rhinoceros (Educational)
Top
Perspective
Front
Right
Layers - All Layers


Command: _Delete
Top
Perspective
Front
Right


Command: _New
Top
Perspective
Front
Right


Loading Rhino Render, version 1.00, Mar 9 2011, 11:41:28
Top
Perspective
Front
Right

The Robotic Platform



From : Bennett & Šabanović (2013), Bennett et al. (2014), Bennett & Šabanović (2015)

What makes a feature “useful” if we are trying to find a pattern?

Feature Must:

1) Contain relevant *information* about target

2) That information is *non-redundant*

1) Feature Selection

- Select a subset of relevant features

2) Feature Extraction (or agglomeration)

- Smush features together

3) Feature Construction (or engineering)

- Create new features out of raw data

Feature Selection – Related Topics

1) Feature Extraction (or agglomeration)

- Dimensionality reduction
- e.g. PCA, Heirarchical Clustering

2) Feature Construction (or engineering)

- Deep Learning
- Manual Feature Engineering

1) Feature Selection

- Select a subset of relevant features

2) Feature Extraction (or agglomeration)

- Smush features together

3) Feature Construction (or engineering)

- Create new features out of raw data

Let’s say you break up with someone, how would you explain it to your friends?

1) Explainability and/or User Adoption

2) Computational Time

3) Signal to Noise, Overfitting, etc.

Feature Selection

1) **Filter Methods**
   - Chi-squared, Gain Ratio, Relief-F, Mutual Information, Low-Variance, Correlation, Regression Based, Symmetrical Uncertainty, etc.

2) **Wrapper Methods**
   - Involves building multiple models on different sets of features, looking for the optimal one
   - Different kinds of search: greedy, random, genetic algorithms

3) **Stepwise Recursive Methods**
   - Stepwise removal, either forward or backward

Feature Selection (cont.)

1) **Filter Methods**
   - ➢Univariate (chi-sq) vs multivariate (relief-f)
   - ➢Target: discrete (gain ratio) vs continuous (mutual info regression)

2) **Wrapper Methods**
   - ➢Embedded methods (feature importance from Random Forests) can be thought of as a "poor man's" approach to this
   - ➢One can create a full-blown wrapper though, encapsulating any kind of ML algorithm (naïve bayes, neural network, etc.)

3) **Stepwise Recursive Methods**
   - ➢More traditional statistical approach

Filter Methods

- **Information Gain (Gain Ratio)**
  - Based on information theory, or *Entropy*

$$H(X) = \sum_{i=1}^{n} p(x_i) I(x_i) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$$

- Gives you a measure of how much "**information**" each feature contains relative to the target
- **Gain Ratio** - solves problem where features with many values or a greater range appear to have more info

Many Filter Methods

- Correlation
- Gini Index
- Chi-squared
- Relief-F
- Symmetrical Uncertainty
- Low-Variance
- etc. etc. etc.

Feature	Gain Ratio
Activity_StrPct	0.1914
ActivityPct	0.1914
Sleepawakenings	0.18425
TotalCaloriesBurned	0.138
UCLAScale	0.04919
WHOQOLDomainAvg	0.04073
Activity_StrCnt	0.01754
ActiveTime	0.01078
ActivityCnt	0.00358

Bayes Net-K2 w/ CAIM discretization

Chi-Sq Value	Variable
31.5054	pre_amt
11.6678	cchg
10.5885	pre_claim_cnt
7.925	age
7.2297	care_program_referral_reason
5.5602	payor
4.7792	arrange_mental_health_visit
4.6721	medical_cost
4.4371	drug_cost
4.1938	overall_cost
4.0159	inperson_contact_cnt
3.5944	prospective_risk
3.2525	arrange_eap
3.2525	avg_duration
3.2241	new
3.1922	first_intervention_type
2.987	failed_contact_cnt
2.5949	encounter_cnt
2.4345	admissions

Wrapper Based Feature Selection

10X Cross-Val (partitioned)					
Model	Binning	# of Features	Accuracy	Correct	Incorrect
Naïve Bayes	CAIM	5	87.7%	157	22
Naïve Bayes	CAIM	8	95.0%	170	9
Naïve Bayes	CAIM	10	96.7%	173	6
Naïve Bayes	CAIM	15	97.8%	175	4
Naïve Bayes	CAIM	All	97.2%	174	5
Naïve Bayes	None	5	86.6%	155	24
Naïve Bayes	None	8	98.3%	176	3
Naïve Bayes	None	10	98.9%	177	2
Naïve Bayes	None	15	98.3%	176	3
Naïve Bayes	None	All	97.8%	175	4
Bayes Net - K2	CAIM	8	95.0%	170	9
Bayes Net - K2	CAIM	10	96.7%	173	6
Bayes Net - K2	CAIM	All	98.3%	176	3
Bayes Net - K2	None	8	92.2%	165	14
Bayes Net - K2	None	10	94.4%	169	10
Bayes Net - K2	None	All	96.7%	173	6
Ensemble	CAIM	8	93.9%	168	11
Ensemble	CAIM	10	96.1%	172	7
Ensemble	CAIM	All	97.8%	175	4
Ensemble	None	8	96.1%	172	7
Ensemble	None	10	96.6%	173	6
Ensemble	None	All	95.5%	171	8

1) Feature Selection

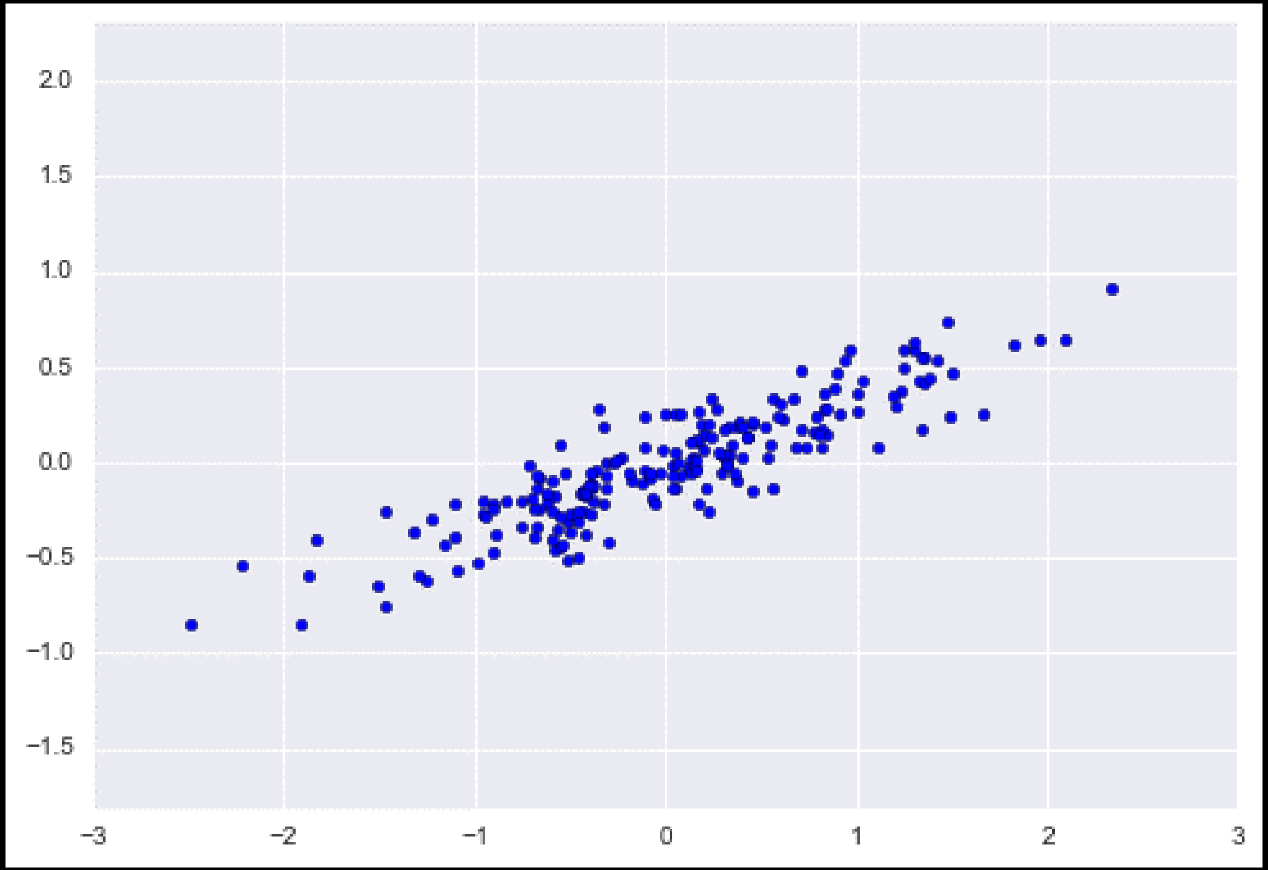
- Select a subset of relevant features

2) Feature Extraction (or agglomeration)

- Smush features together

3) Feature Construction (or engineering)

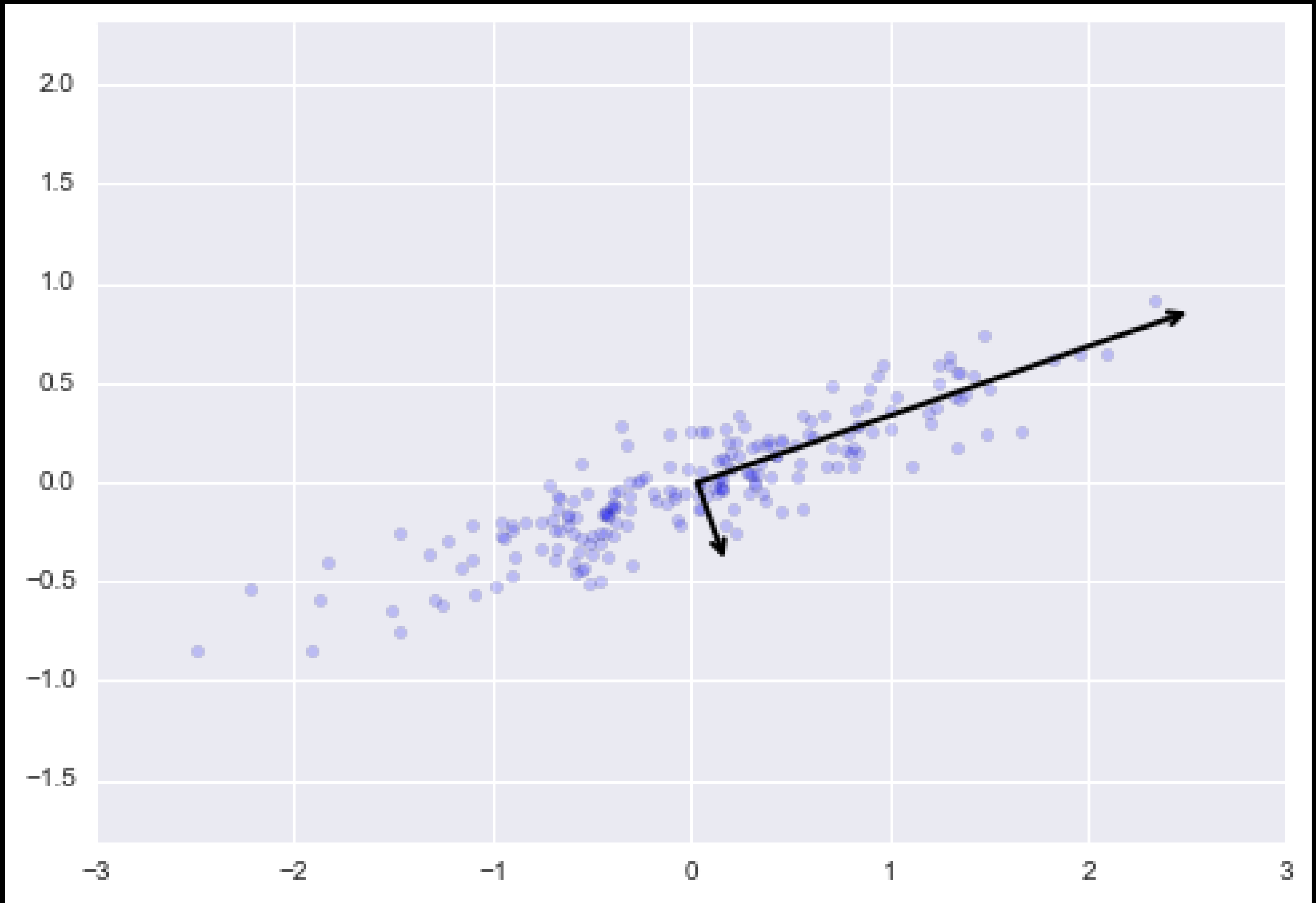
- Create new features out of raw data



Feature 2

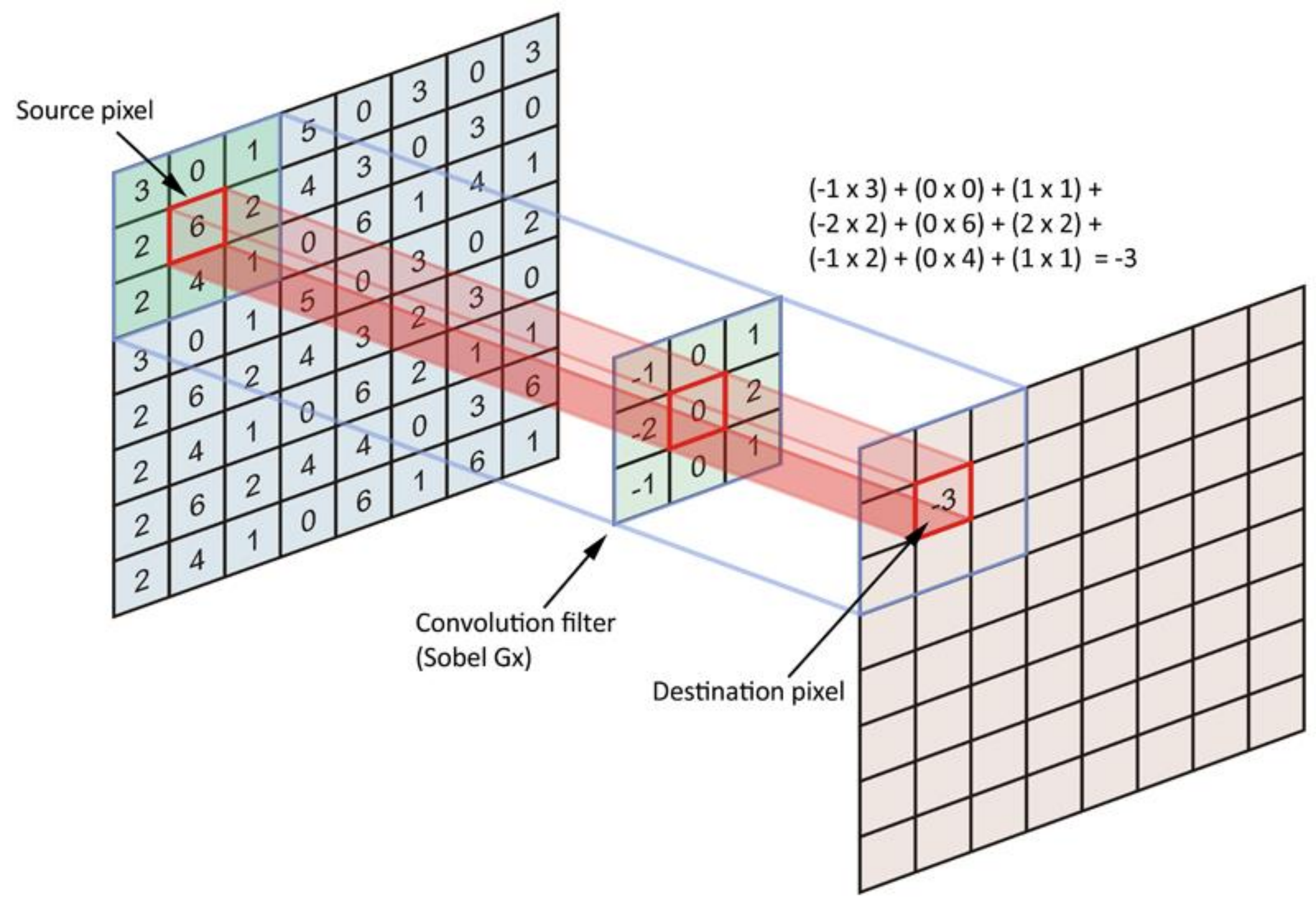
Feature 1

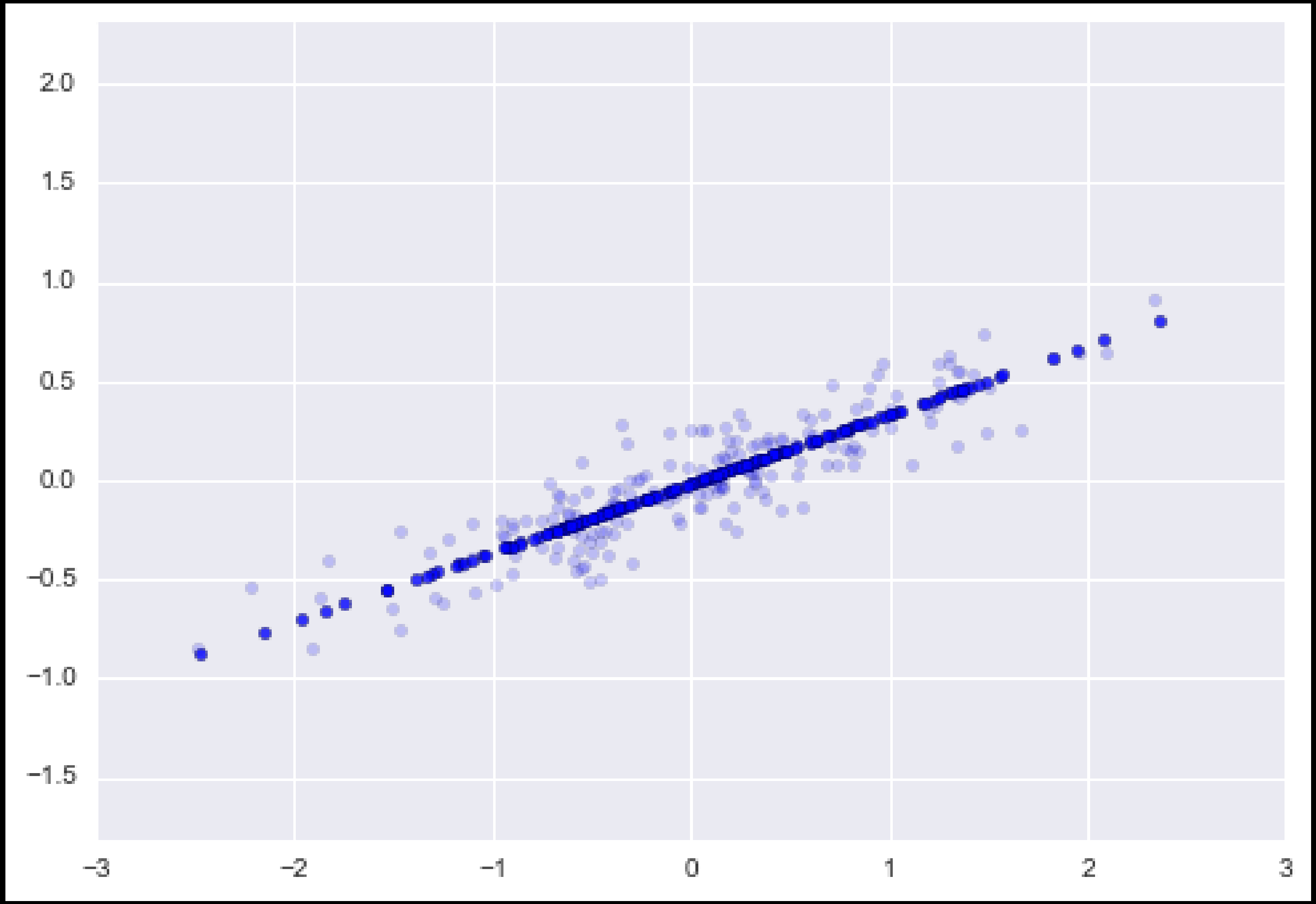
If we have two features, and we want to get rid of one of them, how might we do that?



- Eigenvectors and Eigenvalues of the covariance matrix between features
- **Principle Components**

Convolution Filter





- We can then collapse one of the components into a single dimension (aka “zero it out”)
- **PCA (principle components analysis)**

PCA

- **Can be called in Scikit using:**

```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
pca.fit(X)
```

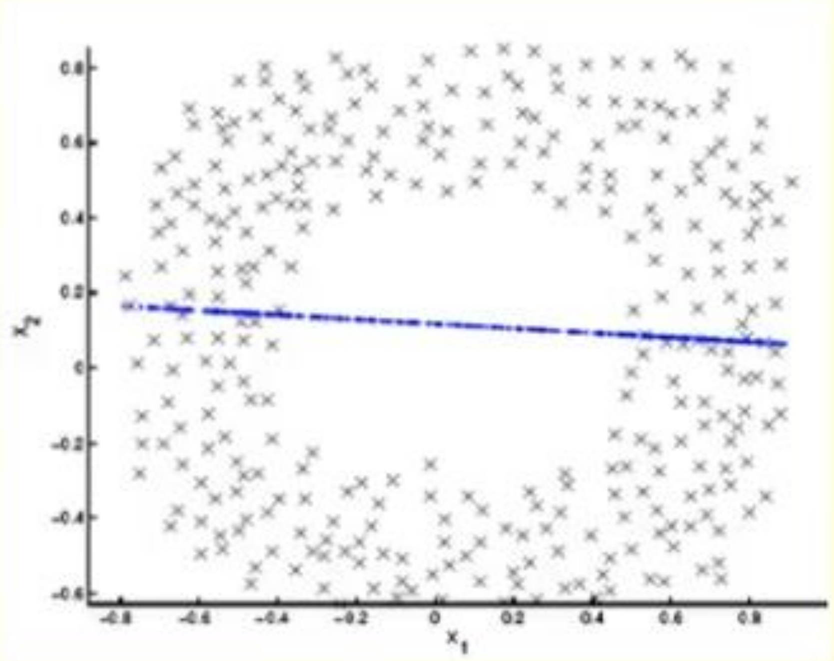
- **In the Spark “linalg” package**

PCA – Main Problem

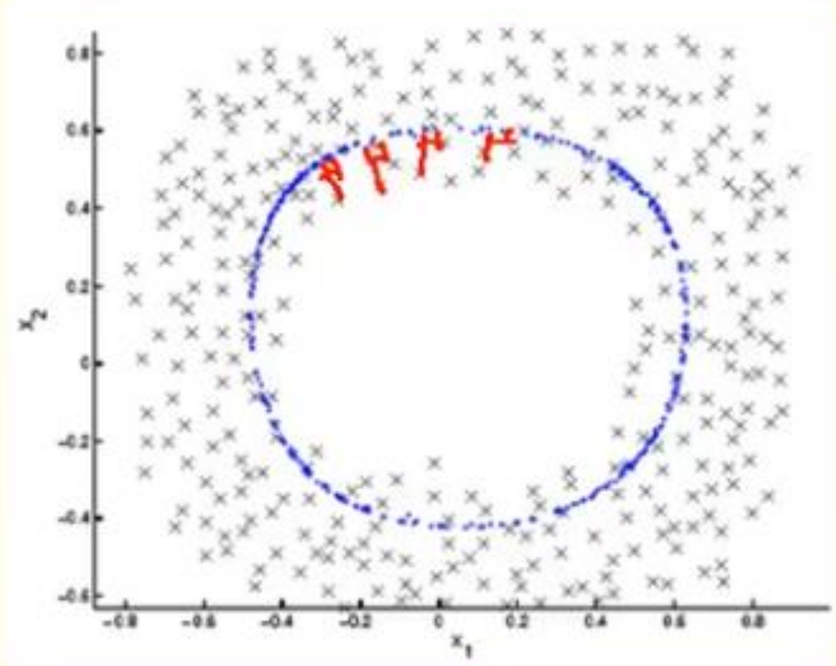
Component Eigenvalues in descending order:

- 1^{st} : 1.910818 (47.8%)
- 2^{nd} : 1.247353 (31.2%)
- 3^{rd} : 0.661220 (16.5%)
- 4^{th} : 0.180607 (4.5%)

Kernel PCA



PCA



KPCA

1) Feature Selection

- Select a subset of relevant features

2) Feature Extraction (or agglomeration)

- Smush features together

3) Feature Construction (or engineering)

- Create new features out of raw data

Feature Construction

Main idea is that we want to create more *relevant* features out of the raw data

1) Manual
   - Domain Experts

2) Automated
   - Deep Learning

Projects

- Read the instructions on the Syllabus closely
- Online students, pay special attention to special instructions in ‘For Online Students’ on D2L
- Presentations due 11/13 and 11/20, assigned *randomly* (see schedule on D2L
- Final Paper due 11/21

Projects

- **Presentation**: Each project is to be presented using PowerPoint in a modified Pecha Kucha style – 20 slides 20 seconds each, on a timer

- Effective Communication – clear succinct, “data science” is your craft

Projects

- **Report**: The report will be written in the format of a paper (abstract, introduction, literature review, methodology, results, discussion, conclusions and future work).
- The literature review for the final report consists of reading and summarizing about 5 to 6 published papers on the review topic. *Proper citations in text*.
- Approximately 6-7 pages long. Single Spaced. Common IEEE conference length.

For next week

1) HW3 Due next week
2) HW4 will release immediately after class that night