CS37300
PURDUE UNIVERSITY

NOV 27 2023

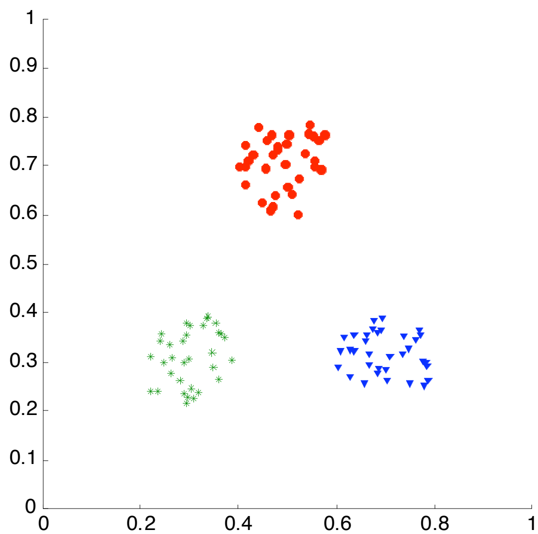
DATA MINING

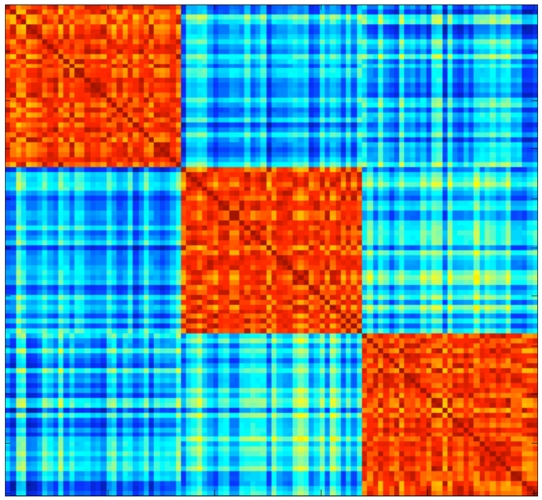
CLUSTERING

- **K-Means, EM**
- **Clustering: Evaluation** – Cohesion and Separation

EXAMPLE 1: GOOD CLUSTERING

Proximity matrix reordered
to reflect cluster assignments





Blue -> Yellow -> Red increasing similarity

CLUSTERING FOR HUMAN UNDERSTANDING (INTERPRETABILITY)

Human beings learn through examples/prototypes, but can over-generalize[1]

Build a mental model of the concept based on prototypes [Newell/Simons 1972]

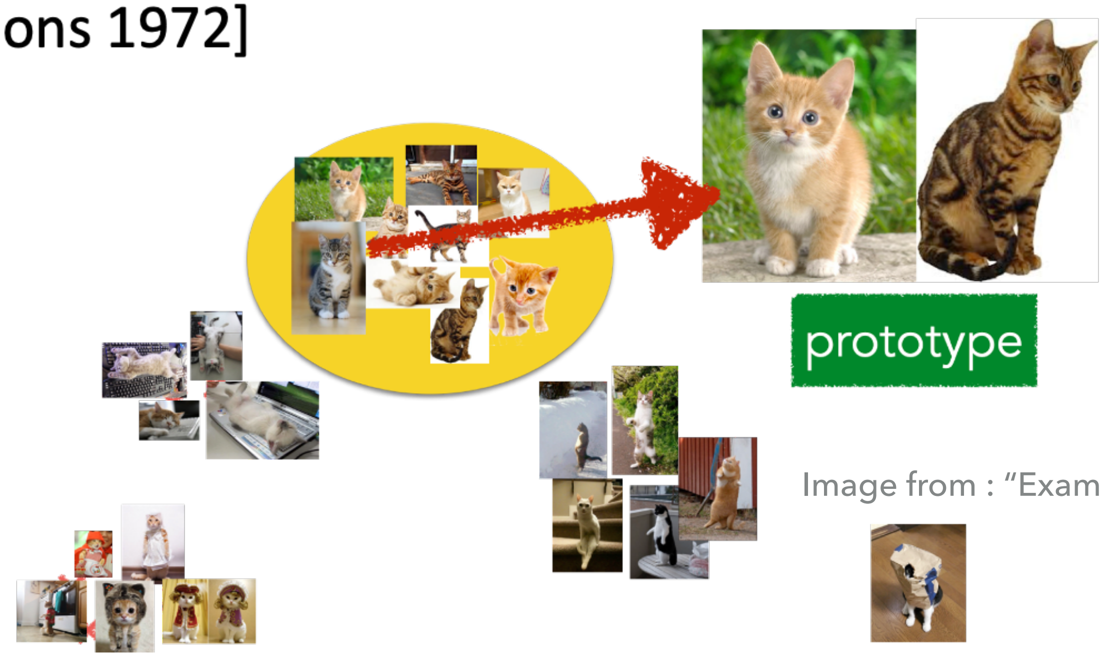


Image from : "Examples are not enough, learn to criticize"

CASE BASE REASONING

- Most of the material here from "Case based reasoning…" Aamodt and Plaza
- Example: Doctor treating similar patients similarly
- Example: Anomaly detection by human beings, such as critical infrastructure inspection, or credit risk assessment
- Example: Business studies
- Condense instances into groups, so as to learn from a case to apply to another similar case
- Applied in various closely-related ways under similar sounding names: Case-based reasoning, examplar based reasoning, instance based reasoning, analogy based reasoning

FOCUS ON HEALTHCARE

- Cluster patients in an unsupervised way, based on their symptoms/readings
- Case based teachings are crucial for multiple reasons:
  - Many diseases not well-understood, not sure what exactly to look for
  - Human body is very complex, we are still uncertain about how various environments, genetics, treatments interact to obtain various outcomes
  - Case-based reasoning can complement more general guidelines for personalized healthcare
  - Example based learning comes naturally to human beings

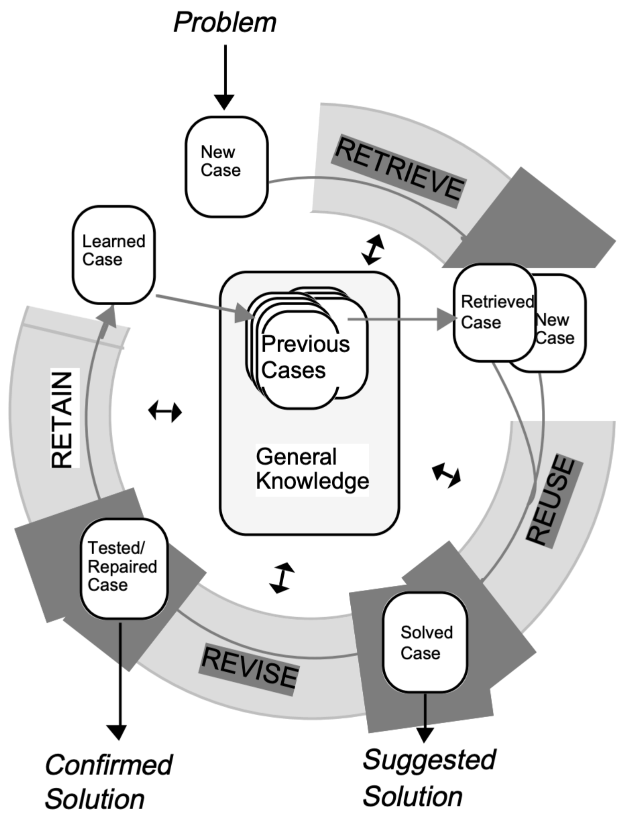


Figure 1. The CBR Cycle

- Image from “Case based reasoning…” Aamodt and Plaza

SIGNIFICANCE AND LIMITATIONS

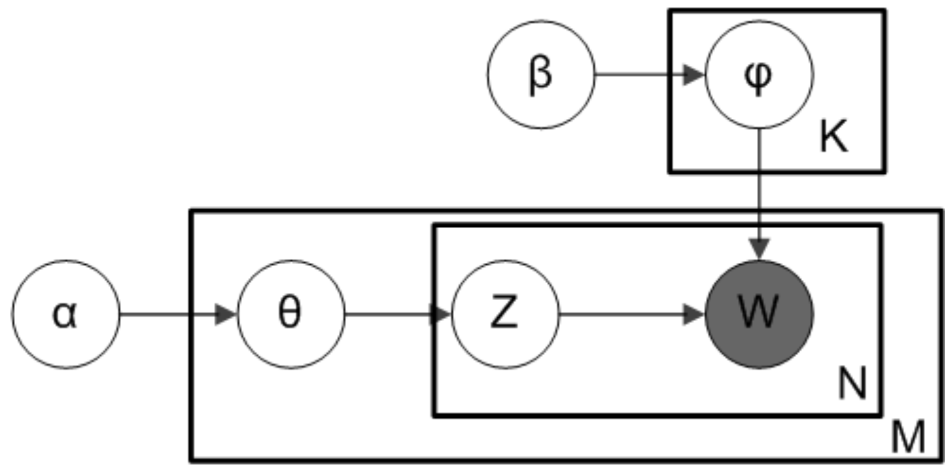
- From “The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification “
- Studies have shown increased human confidence in presence of examples, even more than many competing approaches such as rule-based learning
- Producing explanations is hard – relies on backward engineering of causal relationships (not scalable)
- Increased cognitive load for users for complex similarity measures
- Needs sophisticated expertise for generating explanations.

EXAMPLES OF EASY-TO-EXPLAIN MODELS

- Sparse Linear regression
- Feature-based data split methods such as decision trees/lists
- Example-based methods
  - Nearest neighbors (supervised)
  - Clustering techniques (unsupervised)
  - Topic constructions

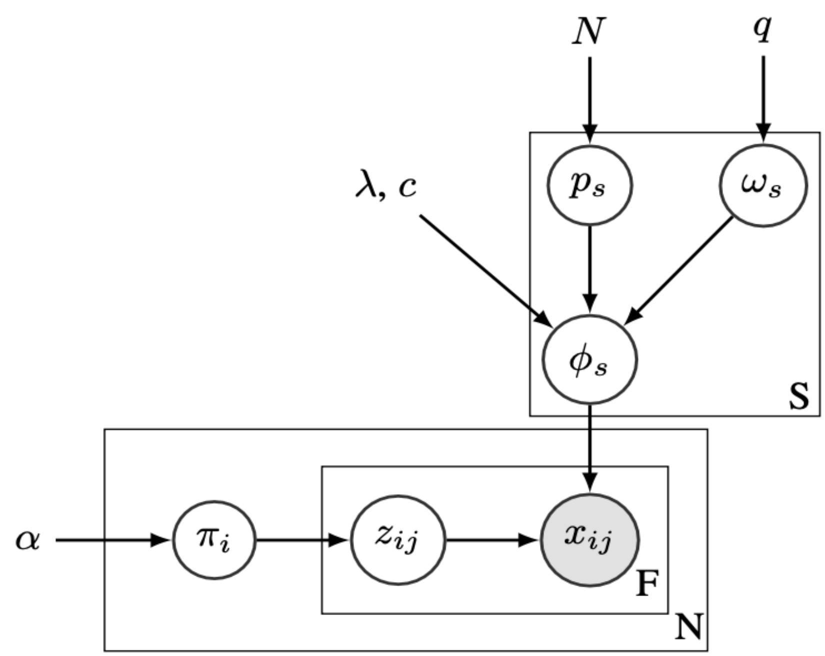
LATENT DIRICHLET ALLOCATION

- Image from wikipedia
- α, β are usually hyper parameters
- θ is topic distribution per doc
- Φ is word distribution per topic
- z is a matrix with topic for every word in every doc
- w is the word matrix indexing every word in every doc



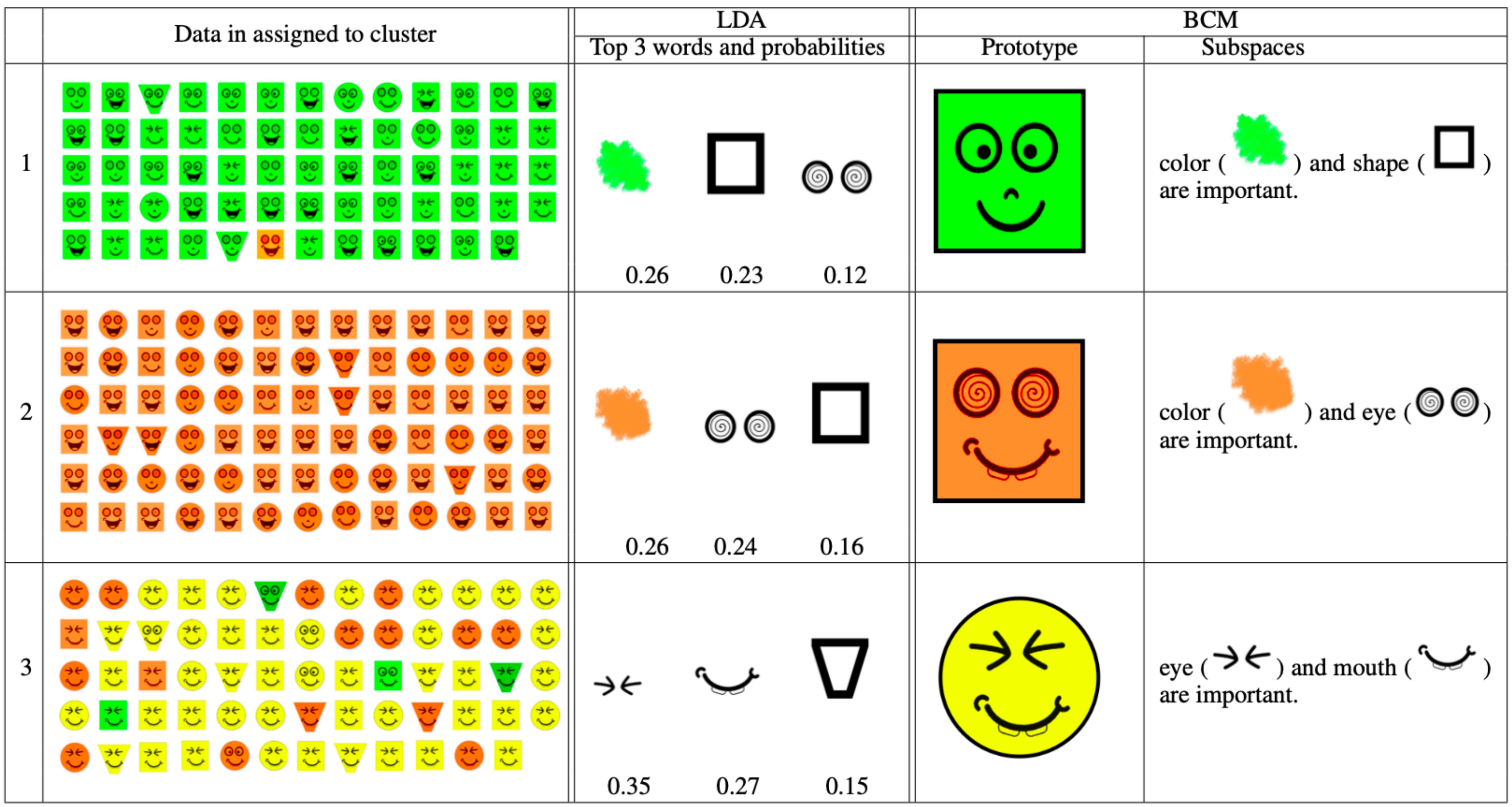
BAYESIAN CASE MODEL

- From “The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification
- Prototypes p_s
- Feature selection \omega_s
- Distribution of feature outcomes \phi_s

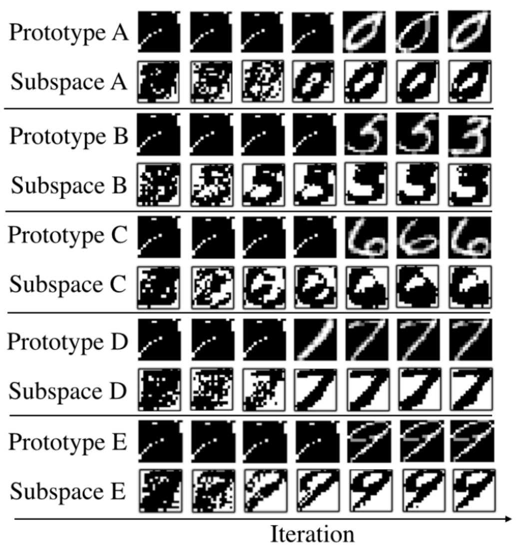


TOPIC SELECTIONS VS PROTOTYPES

- From "The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification



RESULTS



(a) *Handwritten Digit* dataset

Prototype (Recipe names)	Ingredients (Subspaces)
Herbs and Tomato in Pasta	basil, garlic, Italian seasoning, oil pasta pepper salt, tomato
Generic chili recipe	beer chili powder cumin, garlic, meat, oil, onion, pepper, salt, tomato
Microwave brownies	baking powder sugar, butter, chocolate chopped pecans, eggs, flour, salt, vanilla
Spiced-punch	cinnamon stick, lemon juice orange juice pineapple juice sugar, water, whole cloves

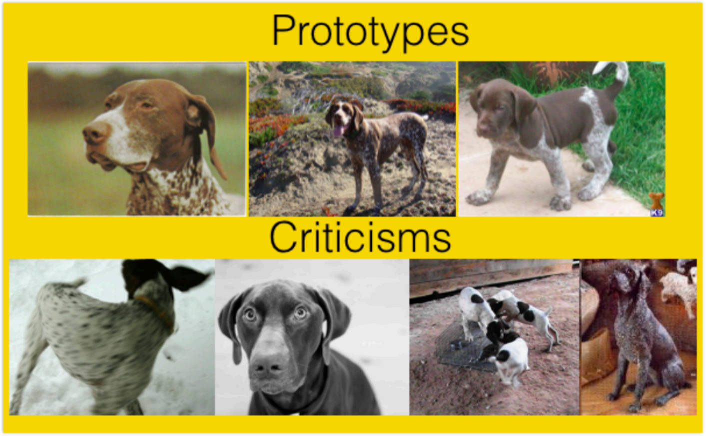
(b) *Recipe* dataset

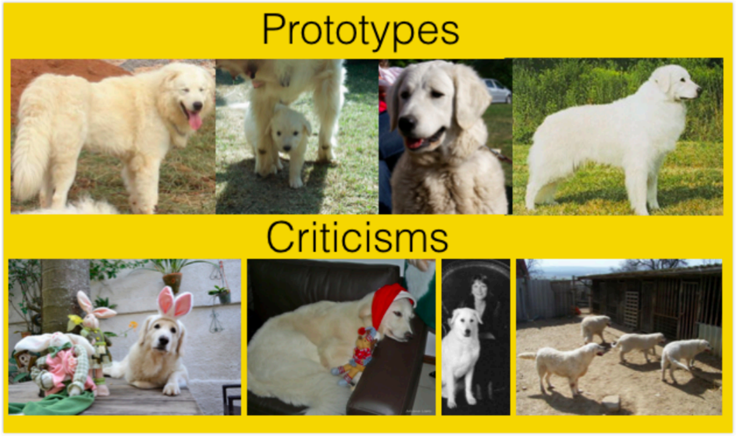
RESULTS — HUMAN EVALUATIONS

- Recipe allocation: Given a set of ingredients only from a recipe, assign them to a cluster using cluster representations learn from LDA and Bayesian Case Model
- BCM explanations are better (85 % accuracy vs LDA 71% statistically significant)

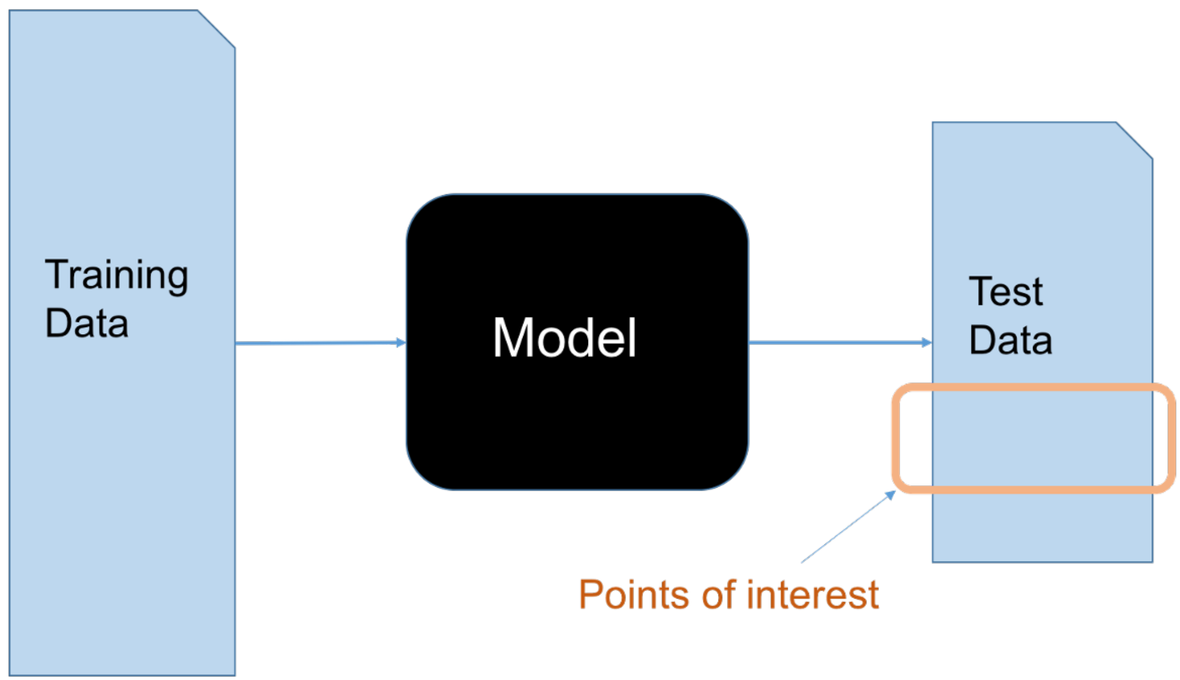
EXAMPLES ARE GREAT, CRITICISMS HELP FURTHER

- From “Examples are not enough, learn to criticize”





- Human experiments – assign images to species (clusters) using prototypes (cluster representatives), random images from clusters, or prototypes+criticisms



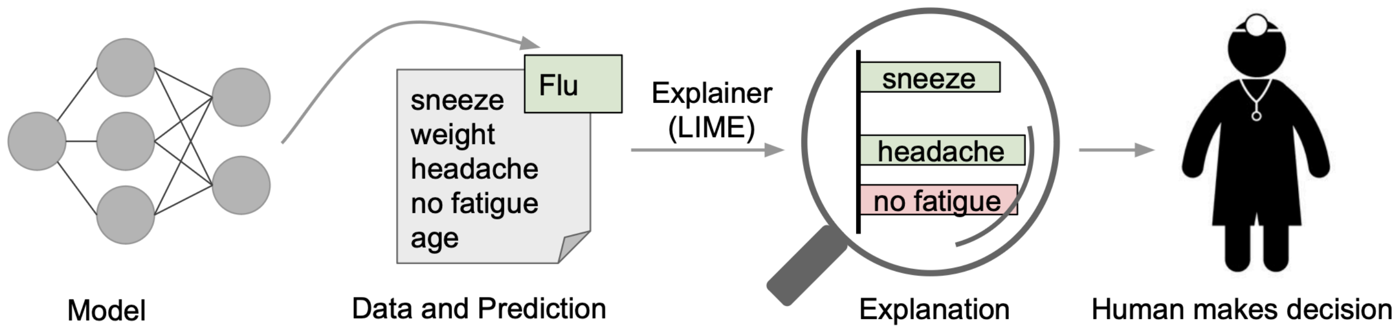
Which training data points are "most" responsible for predicting on points of interest ?

PERFORMANCE ATTRIBUTION

- Feature attribution
  - Linear regression – sparsity
  - Decision trees
- Instance attribution

FEATURE ATTRIBUTION

- From "Why should I trust you?"



DESIGN CRITERIA FOR FEATURE ATTRIBUTION

- Complicated models are hard to understand for laymen
- Choose few features (sparse)
- How do those features impact the prediction?
  - Simpler models are easy to interpret
  - Use simple approximations of the model
- Model agnostic
- High “fidelity”

LIME

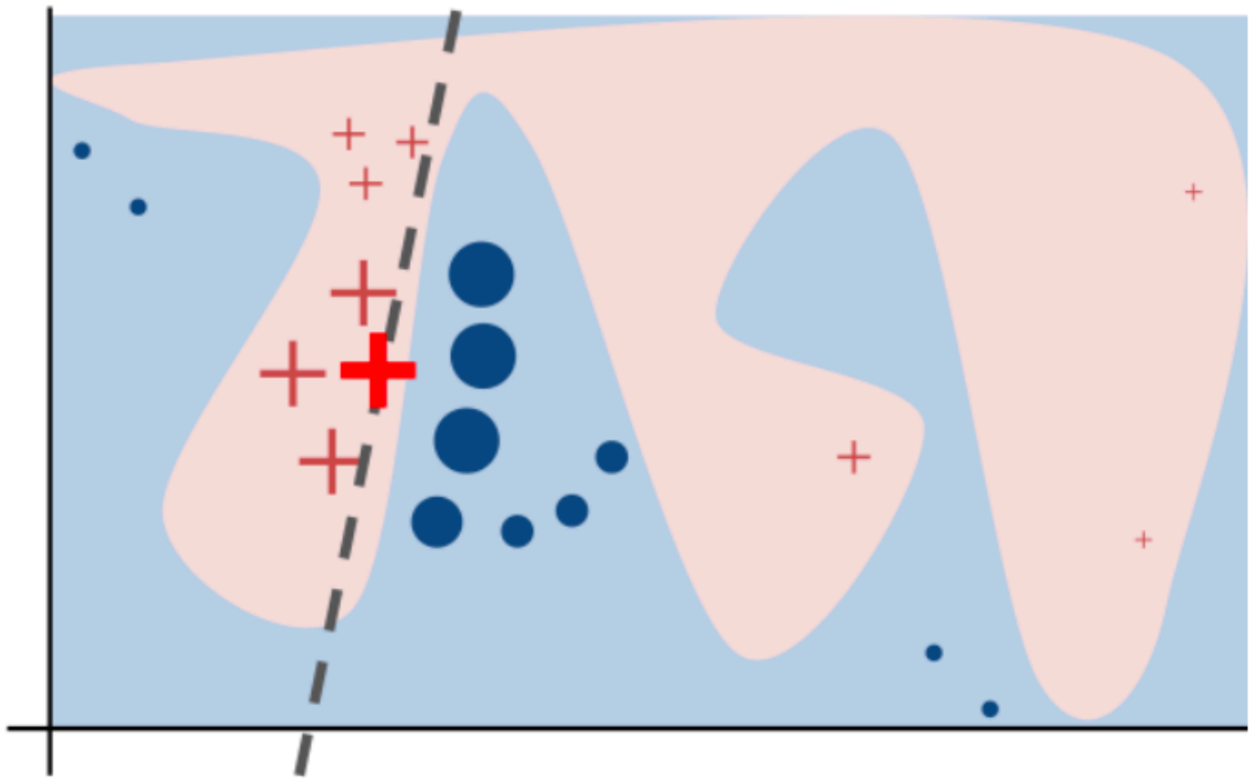
- LIME = Locally Interpretable Model-agnostic Explanations
- Let g() be an easily interpretable simple model
- Let f() be the actual model to interpret
- The idea is to build a g() that closely approximates f() “locally”
-

COST FUNCTION

- Here $\mathscr{L}()$ is the loss that approximates f() with g() in the local region π_x around the instance in question x
- Even though f() may be very complex *globally,* it may still be possible to approximate it *locally.*
- Ω is the complexity measure

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \; \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

EXAMPLE



EXAMPLE

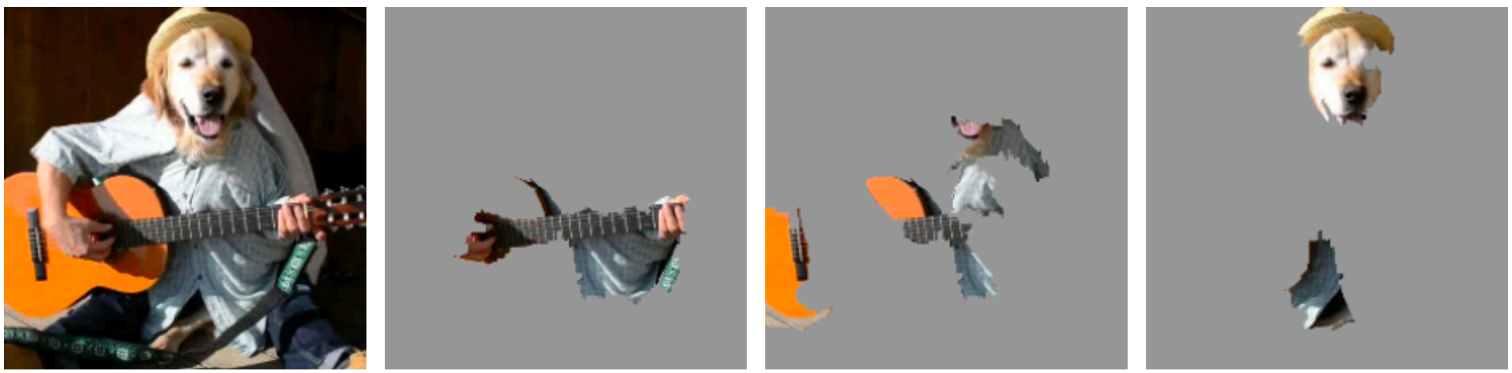
- Let g() be linear functions
- Let $\pi_x := \exp(-D(x,z)/\sigma^2)$ be an exponential kernel for some distance function D
- Loss: $\mathscr{L} = \sum_{z,z' \in Z} \pi_x(z)(f(z) - g(z'))^2$

EXAMPLE (CONTD)

- Ω is measure of complexity *acceptable to the end user*
  - For text: this is could be "use up to k words"
  - For images it could be "use up to k super pixels"
  -

IMAGE EXPLANATION

- Segment the image into super pixels which are x
- Build simple linear model around vicinity of each super pixel to obtain high local fidelity



(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

- **Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)**