

A photograph of a Zen garden featuring a large, circular area of white gravel with concentric ripples. In the background, there are dark, mossy rocks and green foliage.

**Simplicity**



# Linear Functions



- Simple function

$$f(x) = w^\top x + b = \sum_{i=1}^n w_i x_i + b$$

- Easy to interpret everything
  - Weights  $w_i$  tell us how important feature  $i$  is
  - Covariates  $x_i$  tell us how strong this is for specific  $x$
- Problems if we have too many coordinates (e.g. images, microarray data, time series)

# Linear Functions

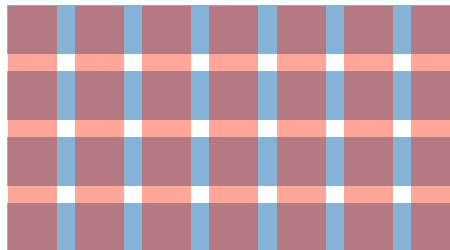


- Simple function

$$f(x) = w^\top x + b = \sum_{i=1}^n w_i x_i + b$$

- Sparsity regularization for extra simplicity
  - Vanilla penalty  $\lambda \|w\|_1$  (with lots of theory for it)
  - Structured penalty (with even more theory for it)

$$\sum_i \|W_{[:,i]}\|_1 + \|W_{[i,:]}\|_1$$



Group penalty

# Linear Functions



- For high-dimensional data this is still too complex
- Linearity is not always a good model
- **Often not suitable for humans**
  - Too tedious to evaluate
  - Difficult to understand
  - Difficult to operationalize **in a stressful situation**

# PADI

## Dive Chart Solves PDE.

## Use while Diving.

**START**

Oxygen p.p. (ata)

DEPTH (feet)

PRESSURE GROUP

45

NO DECOMPRESSION LIMITS

SAFETY STOP REQUIRED

DEPTH SHOWN FOR CONTINGENCY PLANNING ONLY

TABLE 1  
NO DECOMPRESSION LIMITS AND GROUP DESIGNATION TABLE

For planning dives with EANx 32 only by certified enriched air divers.



**PADI**  
padi.com

DISTRIBUTED BY  
PADI AMERICA, INC.

START	0.76	0.80	0.85	0.90	1.00	1.10	1.19	1.29	1.39	1.48	1.58	
DEPTH (feet)	45	50	55	60	70	80	90	100	110	120	130	
A	10	9	8	7	6	5	5	4	4	3	3	
B	20	17	15	14	11	10	8	7	7	6	6	
C	26	23	20	18	15	13	11	10	9	8	7	
D	30	26	23	20	17	14	13	11	10	9	8	
E	34	29	26	23	19	16	14	12	11	10	9	
F	37	32	28	25	21	18	15	14	12	11	10	
G	41	36	31	28	23	19	17	15	13	12	11	
H	46	39	34	30	25	21	18	16	14	13	12	
I	50	43	37	33	27	23	20	17	16	14	13	
J	55	47	41	36	29	25	21	19	17	15	14	
K	60	51	44	39	32	27	23	20	18	16	15	
L	65	55	47	42	34	28	24	22	19	17	16	
M	71	59	51	45	36	30	26	23	20	18	17	
N	77	64	55	48	39	32	28	24	22	19	18	
O	83	69	59	51	41	34	29	26	23	20		
P	90	74	63	55	44	36	31	27	24			
Q	98	80	67	58	46	38	33	29	25			
R	106	85	72	62	49	41	34	30				
S	115	92	77	66	52	43	35					
T	126	99	82	70	55	45						
U	138	106	87	74	58							
V	151	114	93	79	60							
W	167	123	99	84								
X	187	133	106	90								
Y	213	145	110									
Z	220	155										

## RECREATIONAL DIVE PLANNER™

DIVING SCIENCE & TECHNOLOGY, CORP.

TABLE 2  
SURFACE INTERVAL  
CREDIT TABLE

A	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
B	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
C	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
D	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
E	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
F	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
G	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
H	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
I	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
J	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
K	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
L	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
M	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
N	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
O	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
P	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
Q	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
R	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
S	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
T	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
U	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
V	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
W	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
X	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
Y	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00
Z	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00

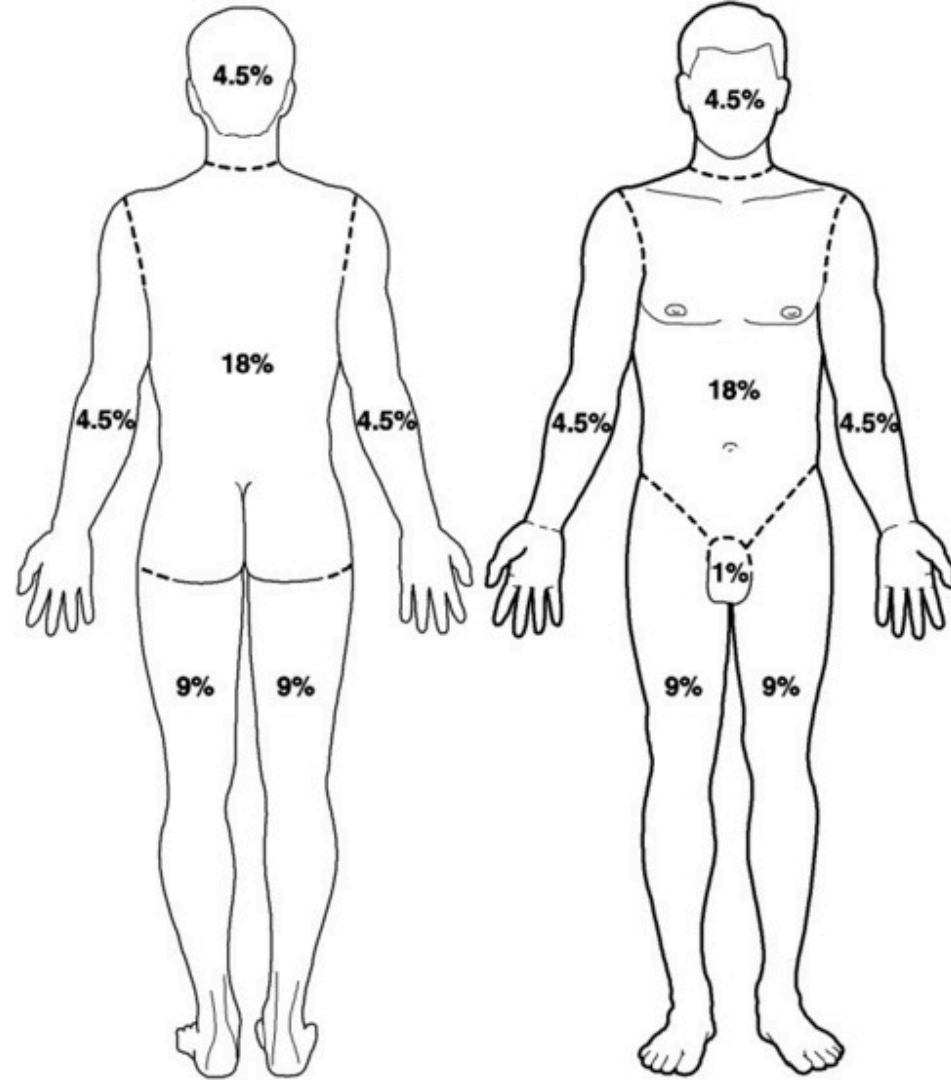
EANx 32  
ONLY

DO NOT USE TO PLAN AIR DIVES

CONTINUE ON OTHER SIDE

# Rule of Nines

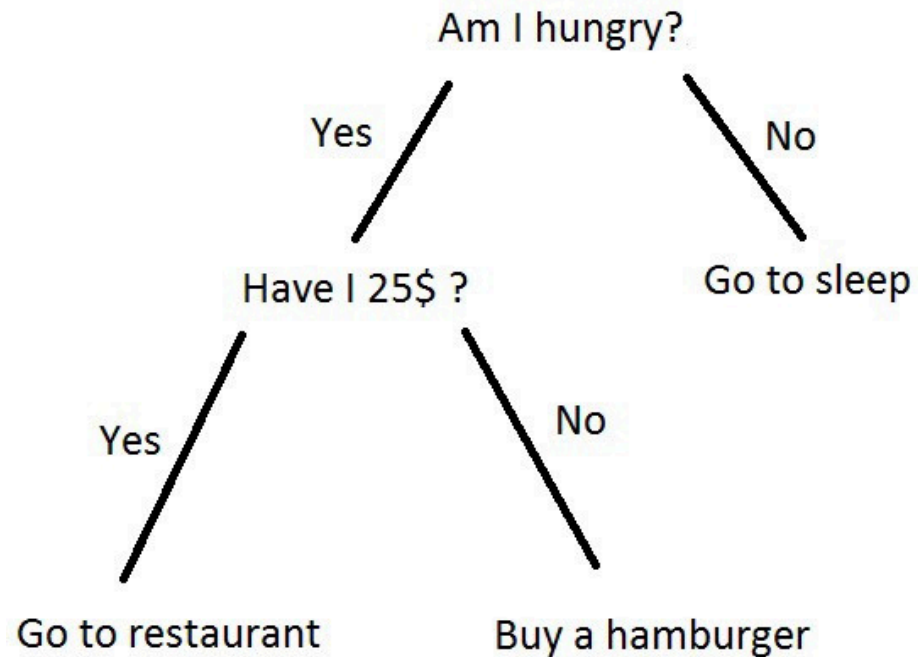
- Trauma Center
  - Assess whether patient will survive their burns.
  - Calculate burned area
  - Use to guide treatment and survival probability.
- Easy to remember (simple percentages)



# Trees and Lists (see e.g. Rudin & coauthors)



- Estimate simple function
  - Decision tree  
(only simple if small)
  - Decision list  
(if then elsif then elsif then)



Type/probability	# bits	code
$P(\text{Trek}) = 0.5$	1	1
$P(\text{Specialized}) = 0.25$	2	01
$P(\text{Cervelo}) = 0.125$	3	001
$P(\text{Serrota}) = 0.125$	3	000





# Approximate Simplicity





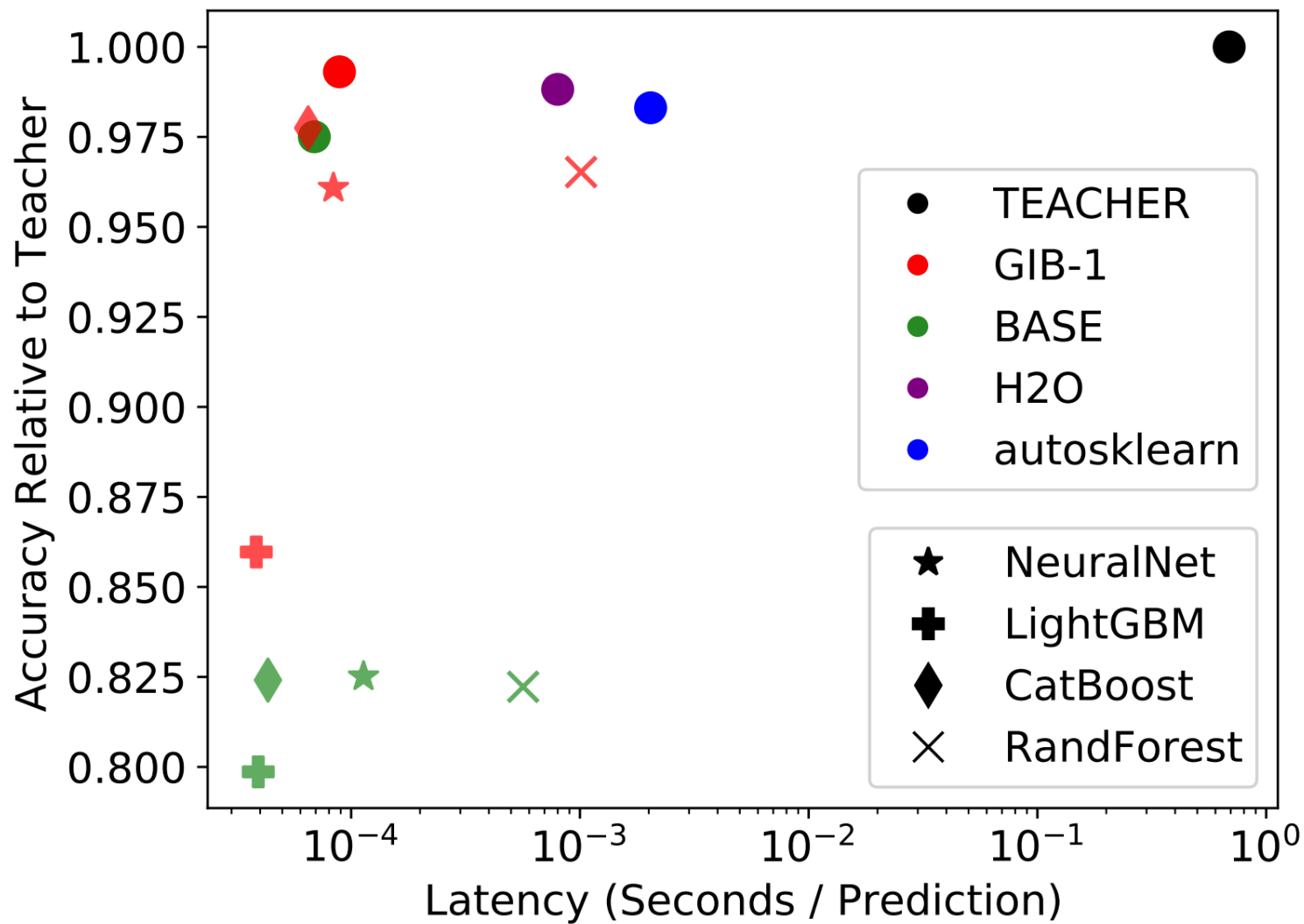
# Alternatives



- Very difficult to estimate efficiently
  - Difficult to incorporate prior knowledge
  - Bagging / stacking / model combinations
  - Large amounts of data
- Distillation to the rescue (train with AutoML then distill)

$$\underset{g}{\text{minimize}} \sum_i l(g(x_i), f(x_i))$$

Fakoor et al., 2020, Fast, Accurate, and Simple Models for Tabular Data via Augmented Distillation, <https://arxiv.org/abs/2006.14284>



# Generating Data for Distillation



- **Training data**

- Might be small
- Already used it to train fancy model  
Labels probably not a lot better than true labels, if at all.

- **Auxiliary data**

- If unlabeled data available, it's perfect
- If not, learn generative model on training data (e.g. via Transformer Density Estimation)
- Gibbs sampler explores neighborhood  $x_i \sim p(x_i | x_{-i})$



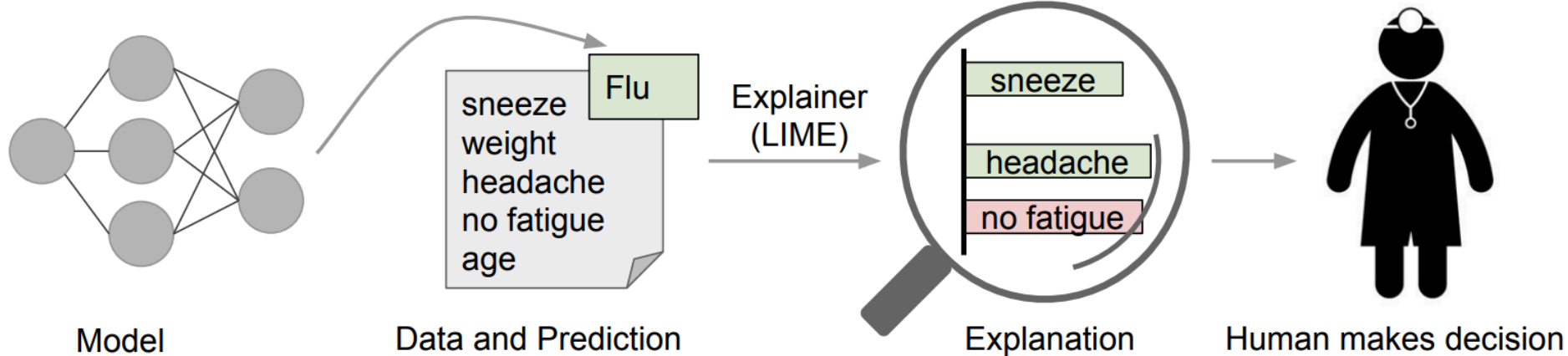
The background of the slide is a close-up, top-down view of numerous sliced lime fruits. The slices are arranged in a dense, overlapping pattern, filling the entire frame. Each slice shows the characteristic green color of the lime pulp, the white pith, and the dark green outer rind. The lighting is bright, highlighting the texture of the fruit segments and the glistening surface of the juice.

# Local Approximation (LIME)

# Generic Classifier



- Black Box classifier can be very complex
- Impossible to approximate well in general
- Maybe can be linearized **locally**



# Local Approximation

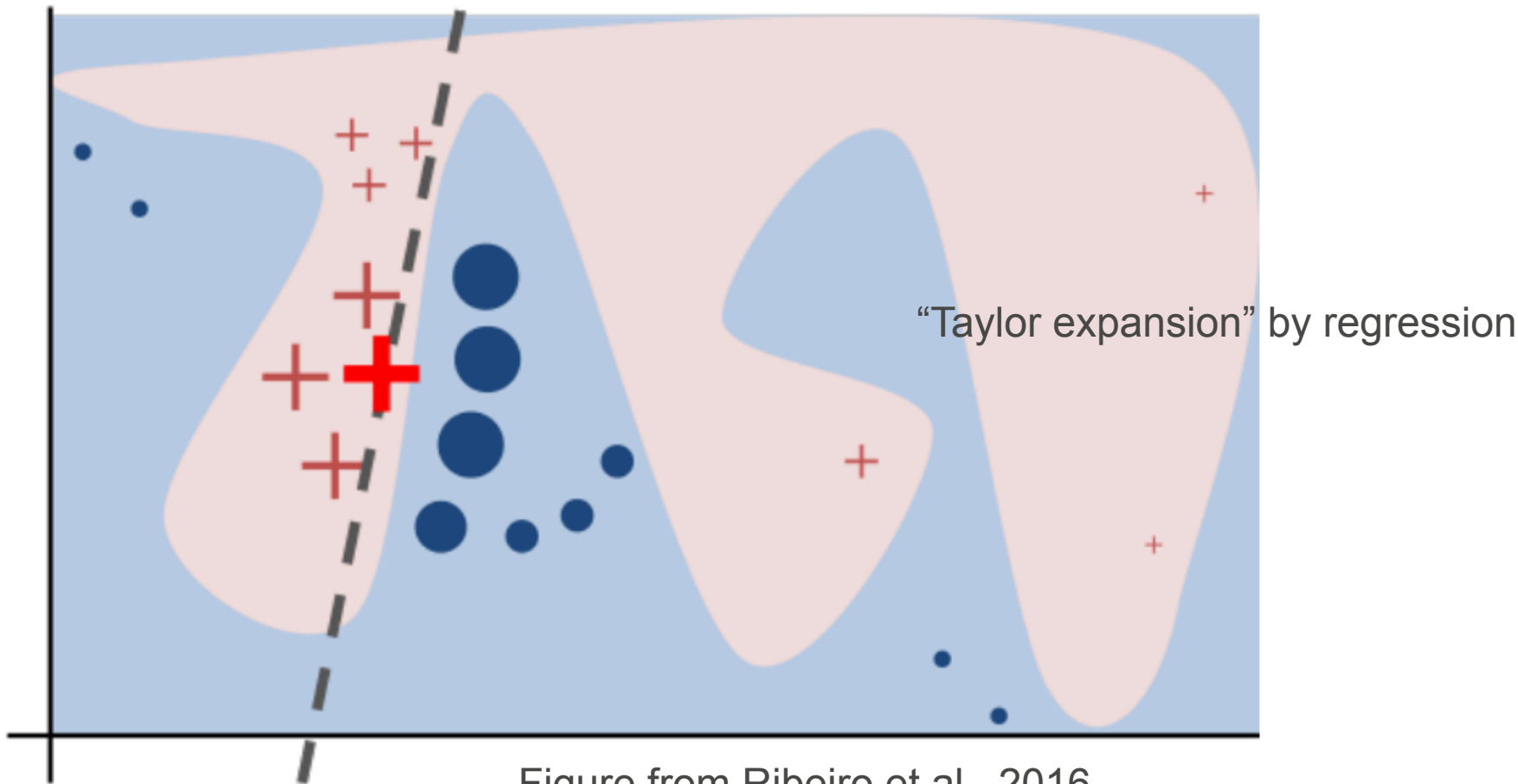


Figure from Ribeiro et al., 2016



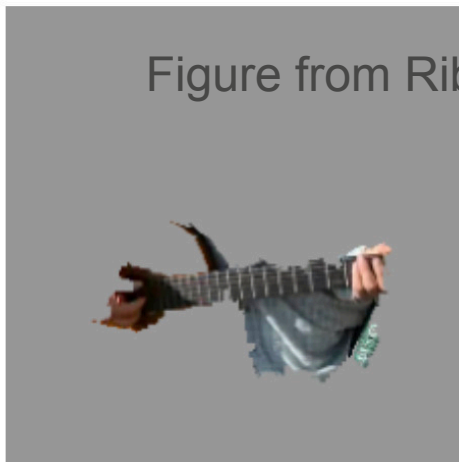
# Locally Linear Approximation



- Get data  $x_j$  around query  $x$
- Approximate  $(x_j, f(x_j))$  pairs with function  $g$
- Can also select across all datapoints for global features (sub modular feature selection)



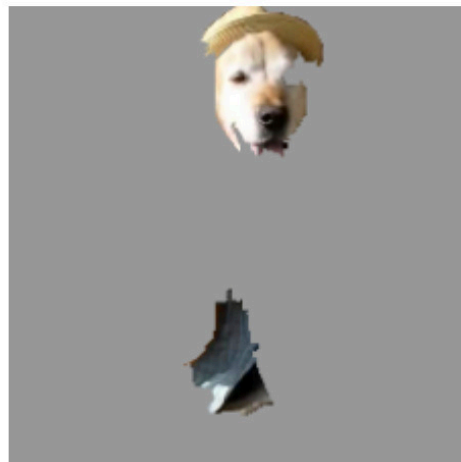
(a) Original Image



(b) Explaining *Electric guitar*



(c) Explaining *Acoustic guitar*



(d) Explaining *Labrador*

Figure from Ribeiro et al., 2016