P²Tab: Prototypical Parts for Tabular Data

Jacek Karolczak† Jerzy Stefanowski

† jacek.karolczak@doctorate.put.poznan.pl

Motivation

"One of the basic characteristics of System 1 is that it represents categories as norms and prototypical exemplars. This is how we think of horses, refrigerators, and New York police officers; we hold in memory a representation of one or more "normal" members of each of these categories."

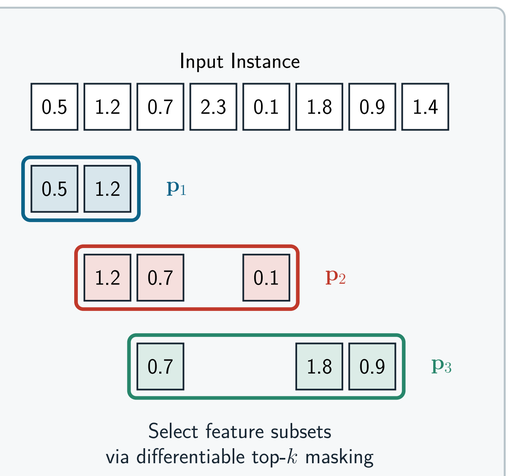
— Daniel Kahneman, *Thinking, Fast and Slow* (2011)

"Categories can be viewed in terms of their clear cases if the perceiver places emphasis on the correlational structure of perceived attributes such that the categories are represented by their most structured portions."

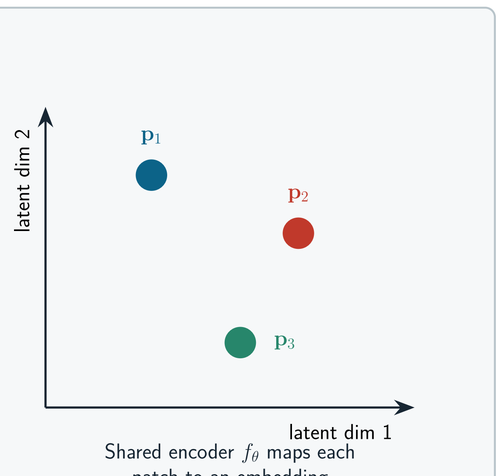
— Eleanor Rosch, *Principles of Categorization* (1978)

P²Tab Step by Step

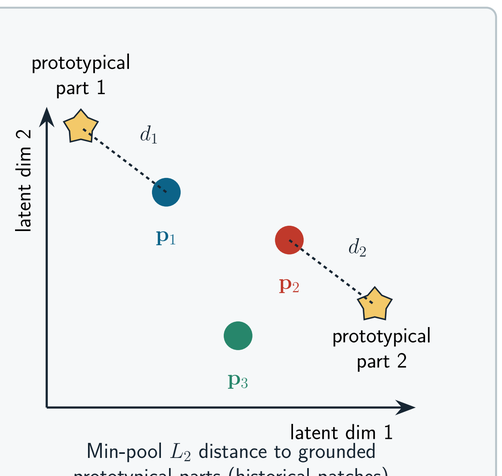
1 Decompose



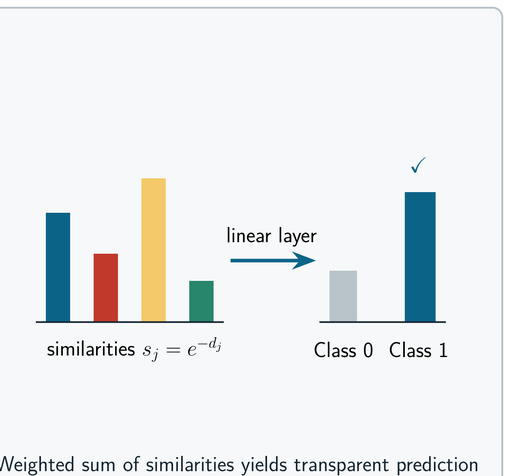
2 Embed



3 Compare



4 Classify



Toy Example

	ExternalRiskEstimate	MSinceOldestTradeO.	AverageMInFile	NumTrades60Ever2D.	NumTrades90Ever2D.	PercentTradesNeverD.	MaxDelq2PublicRecL.	NumTotalTrades	NumTradesOpeninL.	MSinceMostRecentI.	NumInqLast6M	NumInqLast6Mexcl.	NetFractionRevol.	NetFractionInstallB.	NumRevolvingTrades.	NumInstallTrades.	NumBank2NatlT.	PercentTradesWB.	...	L_2 Embeddings Distance **d**
Full Instance	88	-8	129	0	0	100	7	10	1	-7	3	3	4	-8	1	1	0	40	...	
Feature Importance	.69	.38	.83	.31	.47	.36	.51	.82	.96	1.0	.36	.84	.47	.36	.58	.16	.40	.47	...	
Prototypical Part 1	84	–	–	–	–	–	6	–	0	–	–	–	–	79	–	–	–	–	–	.0931
Nearest Instance Patch	88	–	–	–	–	–	7	–	1	–	–	–	–	-8	–	–	–	–	–	
Prototypical Part 2	–	–	–	–	0	–	–	36	–	–	–	–	–	–	0	–	–	60	–	.0977
Nearest Instance Patch	–	–	–	–	0	–	–	10	–	–	–	–	–	–	1	–	–	40	–	
Prototypical Part 3	80	–	–	–	–	–	6	–	1	–	–	–	–	-8	–	–	–	–	–	.1270
Nearest Instance Patch	88	–	–	–	–	–	7	–	1	–	–	–	–	-8	–	–	–	–	–	
...									...											...
Prototypical Part 30	–	211	–	1	1	–	–	–	–	–	–	–	–	–	–	–	–	9	–	.5202
Nearest Instance Patch	–	-8	–	0	0	–	–	–	–	–	–	–	–	–	–	–	–	0	–	
Prototypical Part 31	–	–	36	–	–	–	–	–	3	-7	–	7	–	–	–	–	–	–	–	.5443
Nearest Instance Patch	–	–	129	–	–	–	–	–	–	–	–	–	4	–	–	1	–	40	–	
Prototypical Part 32	–	–	25	–	–	–	–	–	3	2	–	1	–	–	–	–	–	–	–	.6568
Nearest Instance Patch	–	–	–	–	0	100	–	10	–	–	3	–	–	–	–	–	–	–	–	

↓

b + $W \exp(-\mathbf{d}) \rightarrow$ Class 1

A single HELOC instance classified by P²Tab, where each prototypical part is matched to its nearest input patch and the resulting distances drive the prediction.

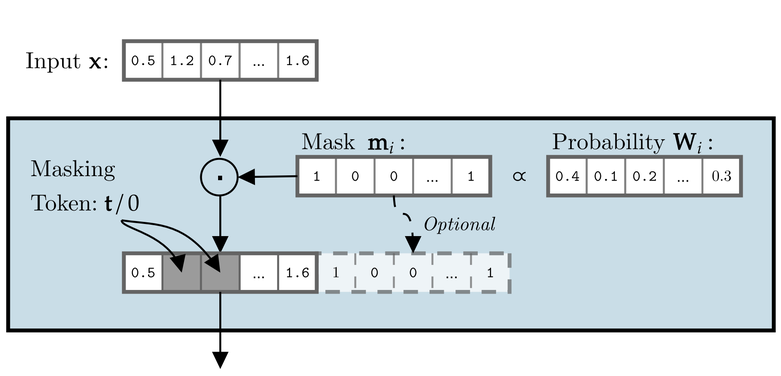
Results

P²Tab matches established baselines with built-in interpretability.

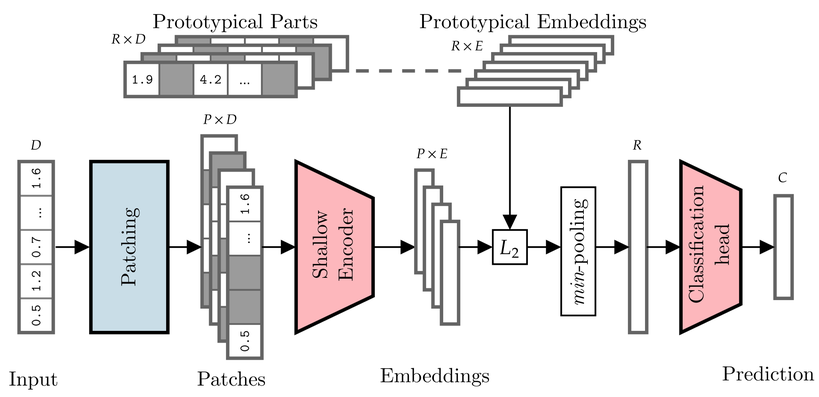
Dataset	Tree	RF	XGB	MLP	P²Tab
CodRNA	.966	.980	**.981**	.972	.967
Credit Card	.658	.652	.655	.653	**.696**
Diabetes	.588	**.712**	**.723**	.644	**.721**
HELOC	.697	.714	**.728**	**.716**	.707
Ionosphere	.810	.907	**.918**	**.973**	**.929**
Pendigits	.888	.937	**.956**	.943	.930
Statlog	**.729**	**.734**	**.729**	.717	**.729**

Balanced Accuracy across 7 datasets.
Bold = not significantly different from the best ($p \geq 0.05$, bootstrap).

Architecture



The patching layer decomposes a tabular instance into granular feature subsets via differentiable top-k masking.



P²Tab maps the input to class probabilities through differentiable patching, latent encoding, and distance-based prototype comparison.

This work is supported by Poznan University of Technology under grant 0311/SBAD/0772.