

In search of Higgs-self-interaction with Machine Learning

B.Sc. Bao Tai, Le
15.1.2025





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

1. Motivation



A place to discover a new theory

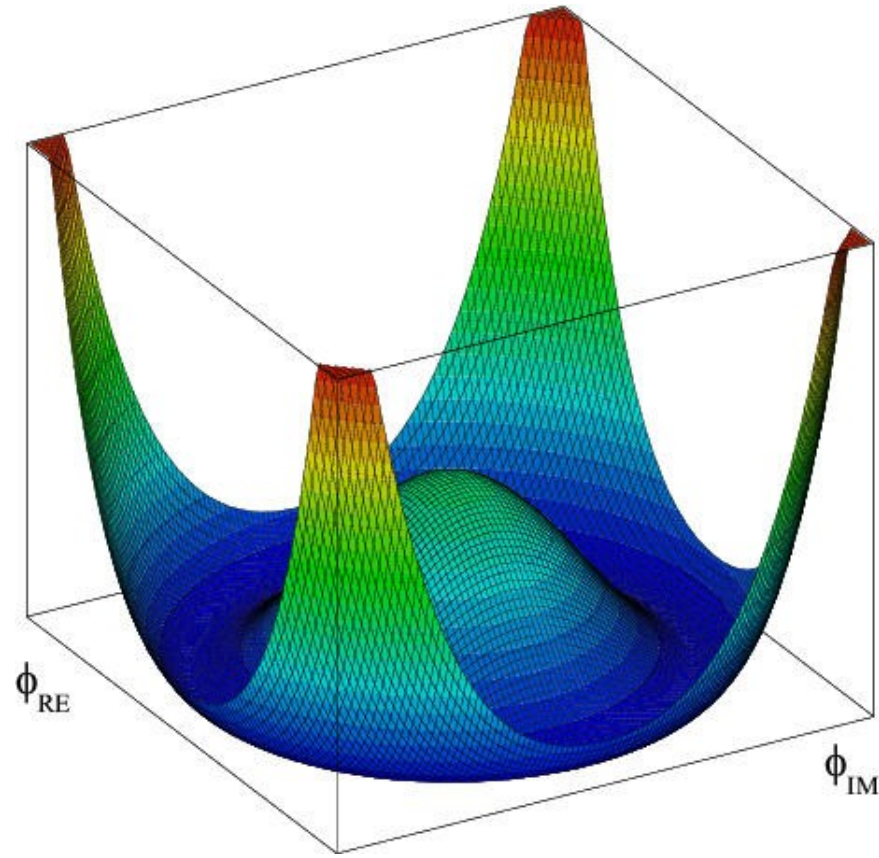
The Nobel Prize in Physics 2013



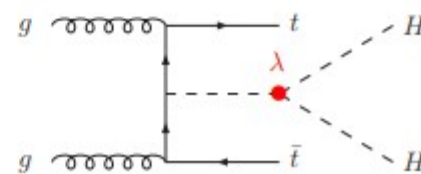
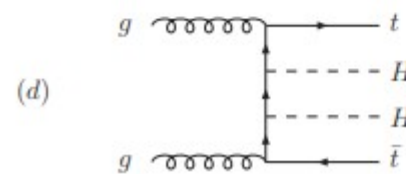
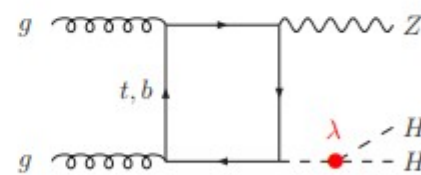
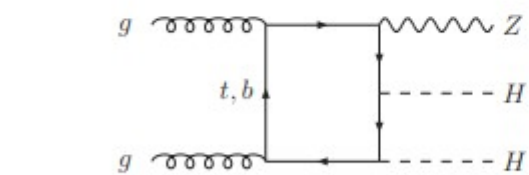
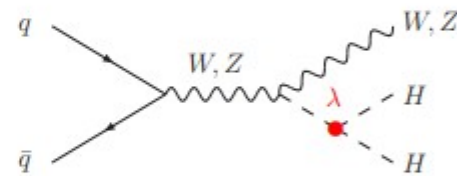
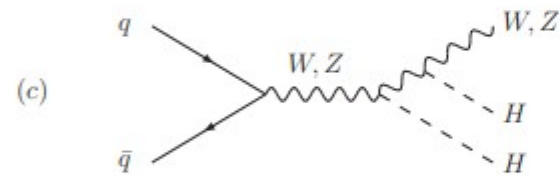
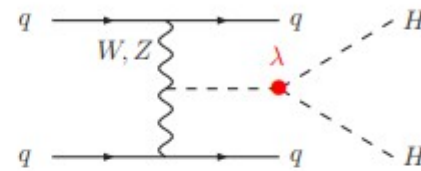
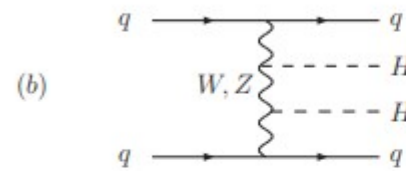
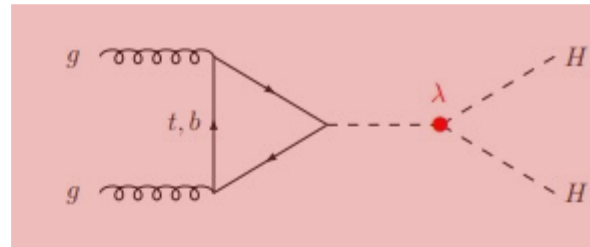
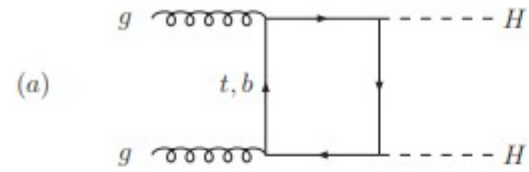
© Nobel Media AB. Photo: A.
Mahmoud
François Englert
Prize share: 1/2



© Nobel Media AB. Photo: A.
Mahmoud
Peter W. Higgs
Prize share: 1/2

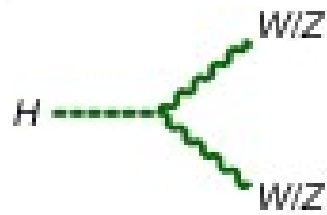


$$V(\phi) = -\frac{\mu^2}{2} |\phi|^2 + \frac{\lambda^2}{4} |\phi|^4$$

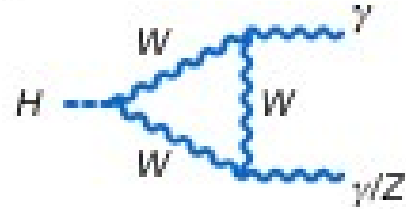


HH \rightarrow 4X

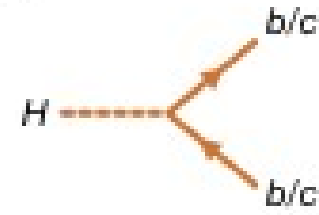
f



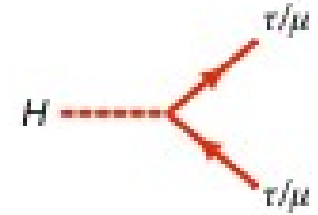
g



h



i





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

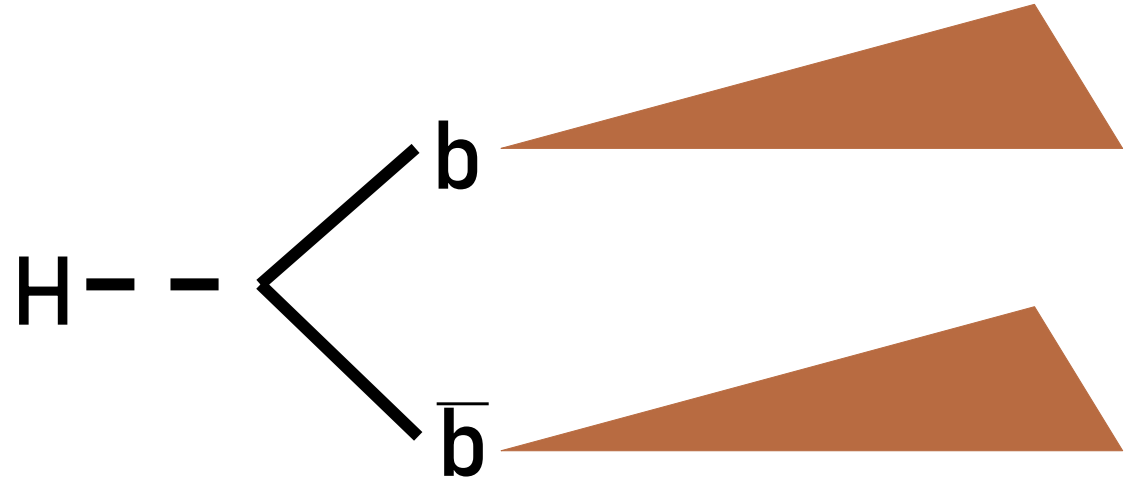
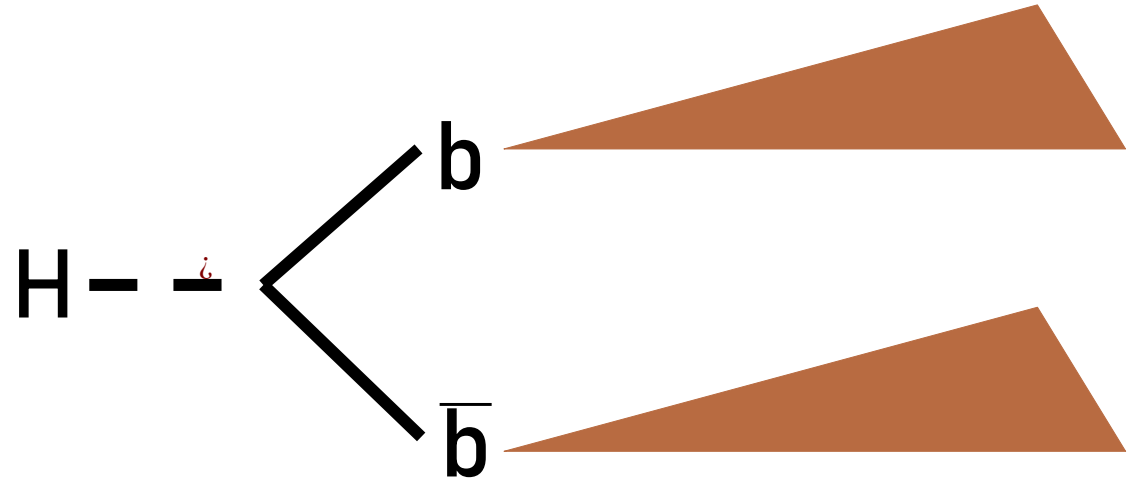


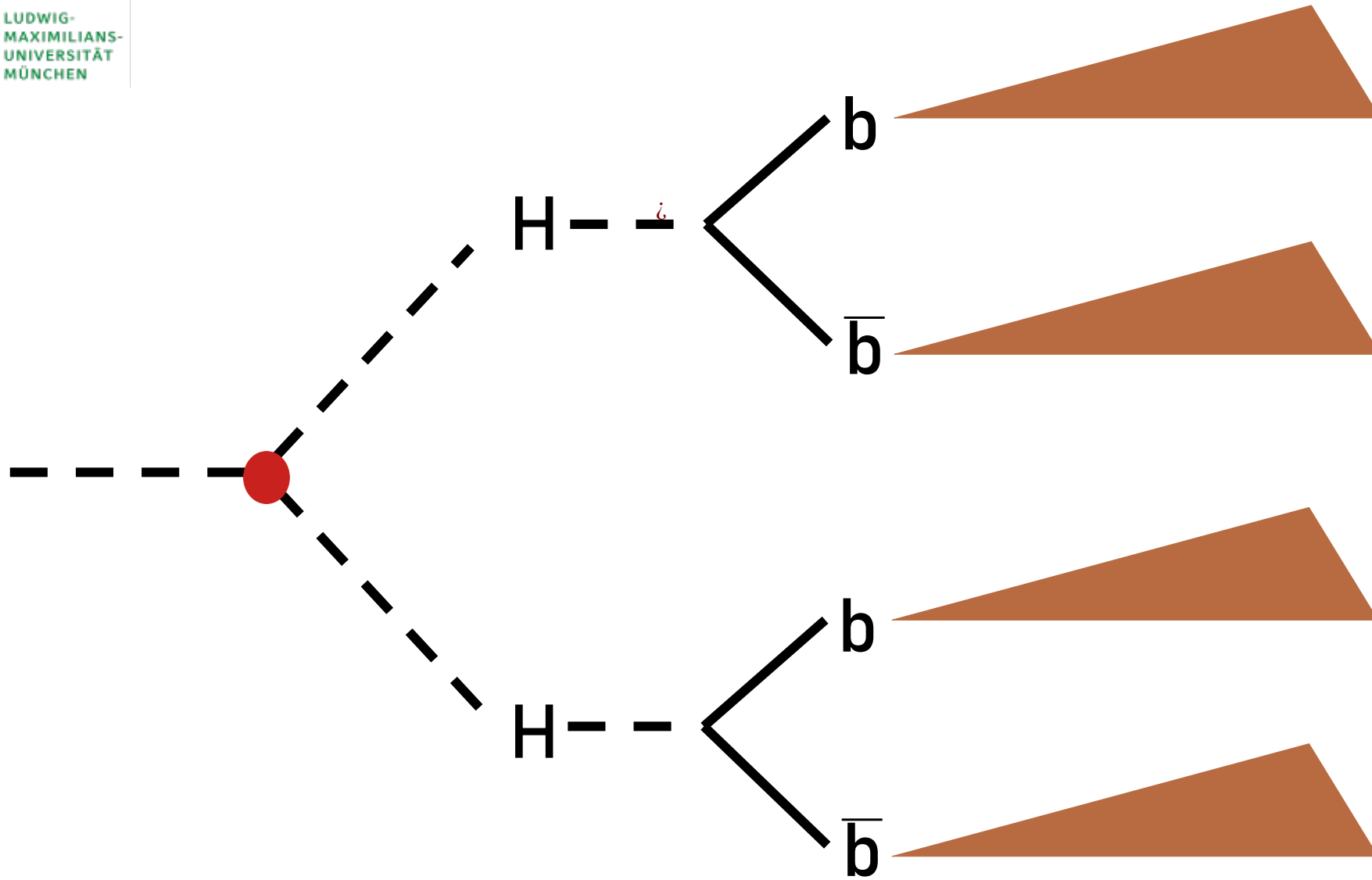


LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

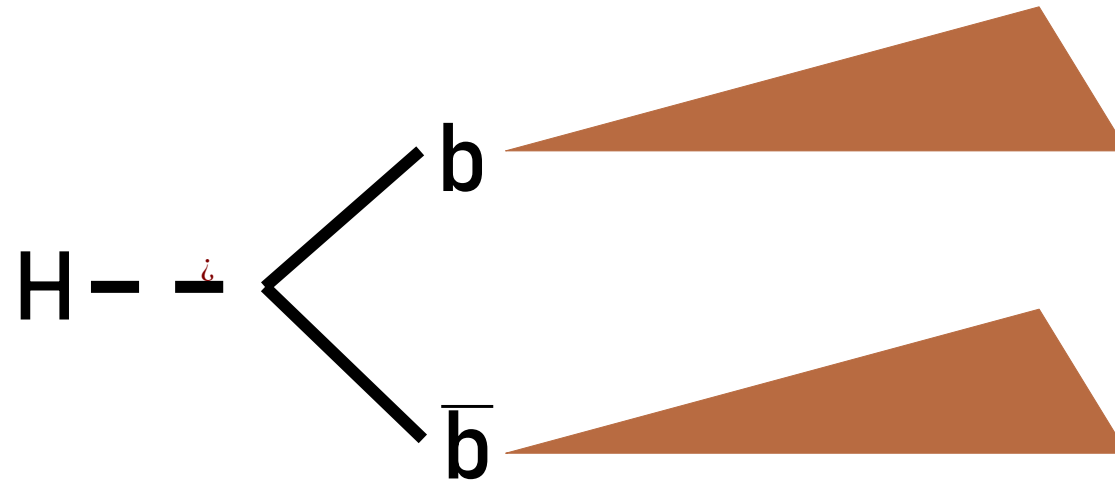
HH → 4X





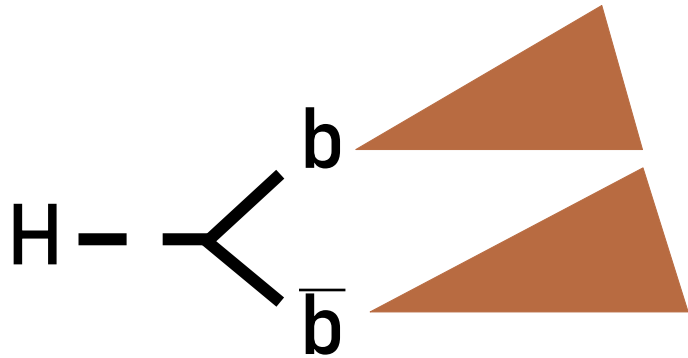
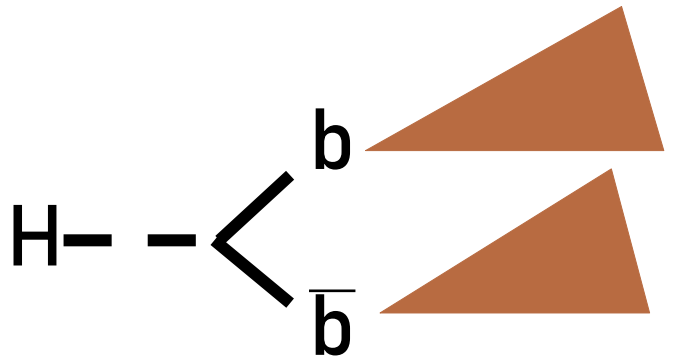


Increasing the HH→4b matching accuracy with Triplet Learning

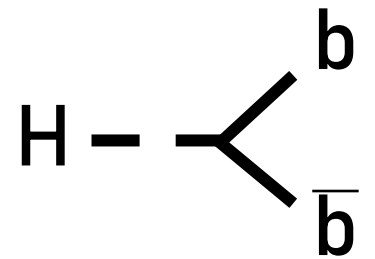
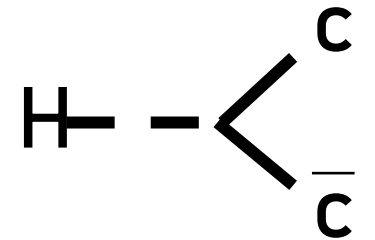
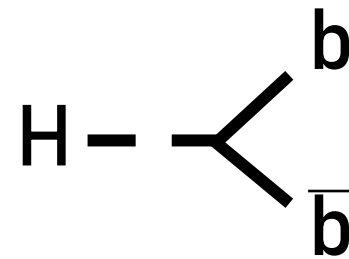
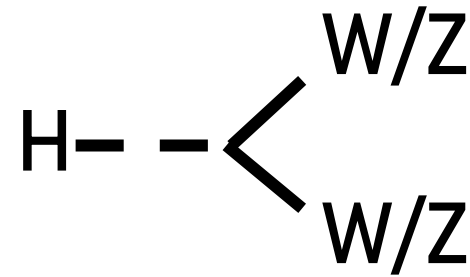


- high branching ratio (58% $H \rightarrow b$ and $H \rightarrow b\bar{b}$ 34%)
- relative to other decay channels higher yield of data
- easy to identify off-vertices of the b-quark decay

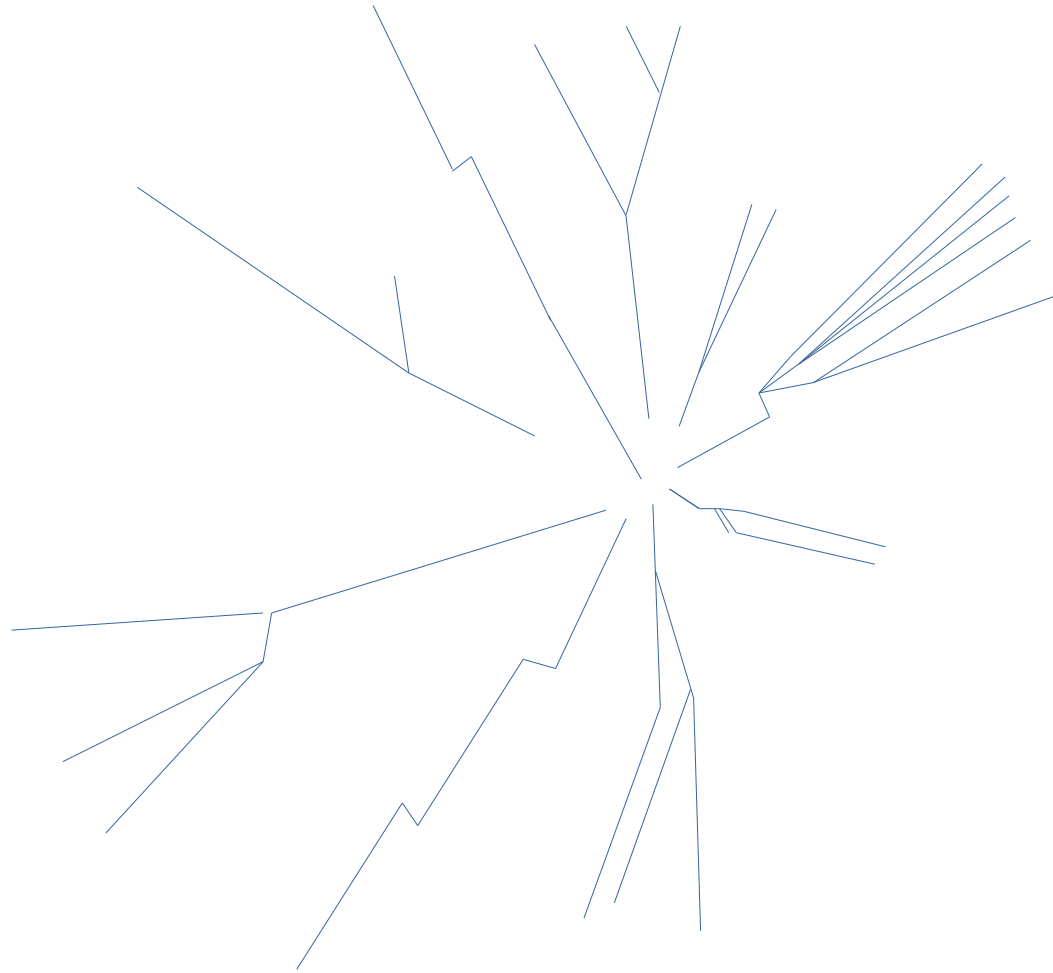
Label
1



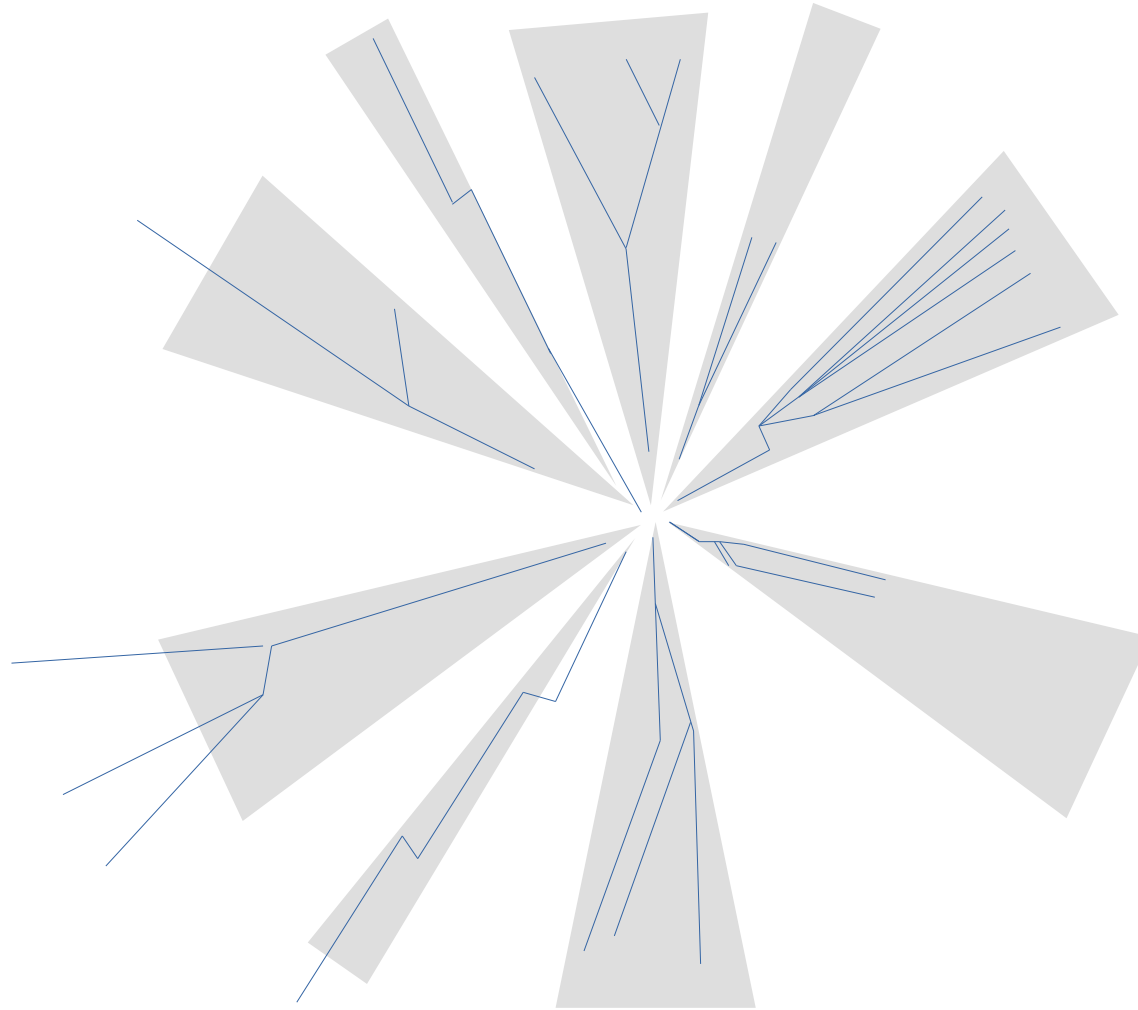
Label
0



- gg → HH Dataset all simulated with PowerHeg and Pythia6



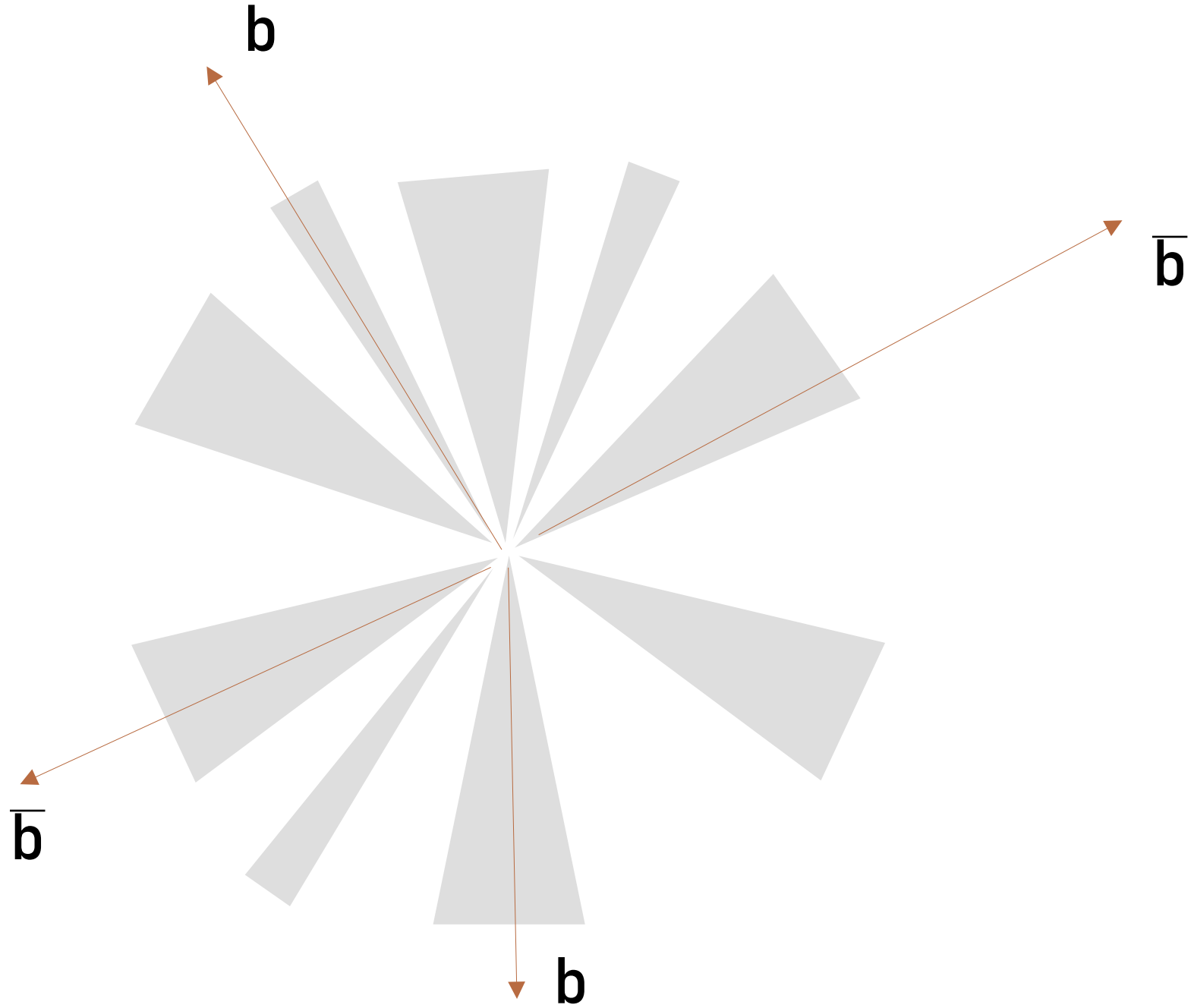
- jets are reconstructed via anti-kt algorithm with a radius of 0.4

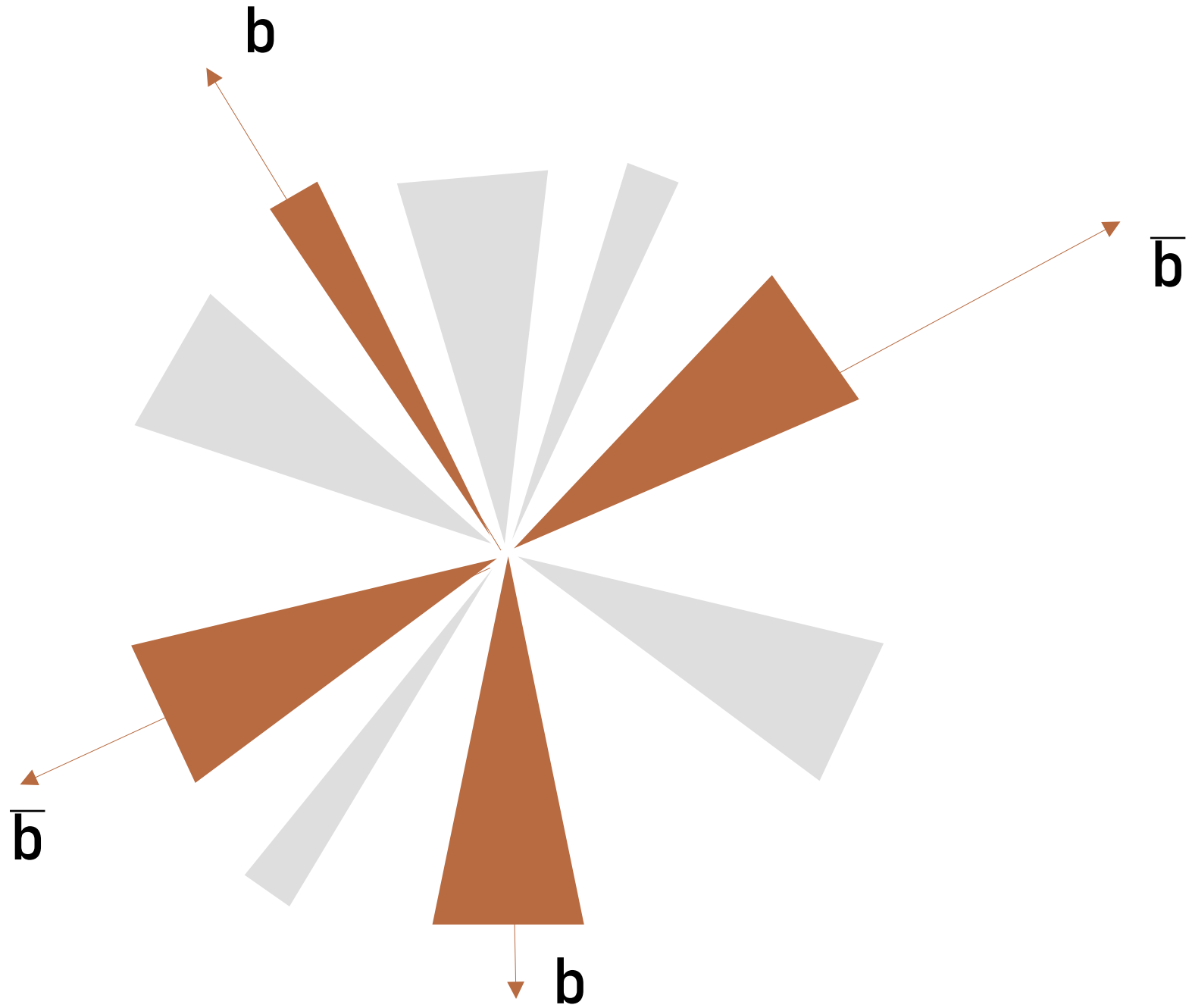




LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

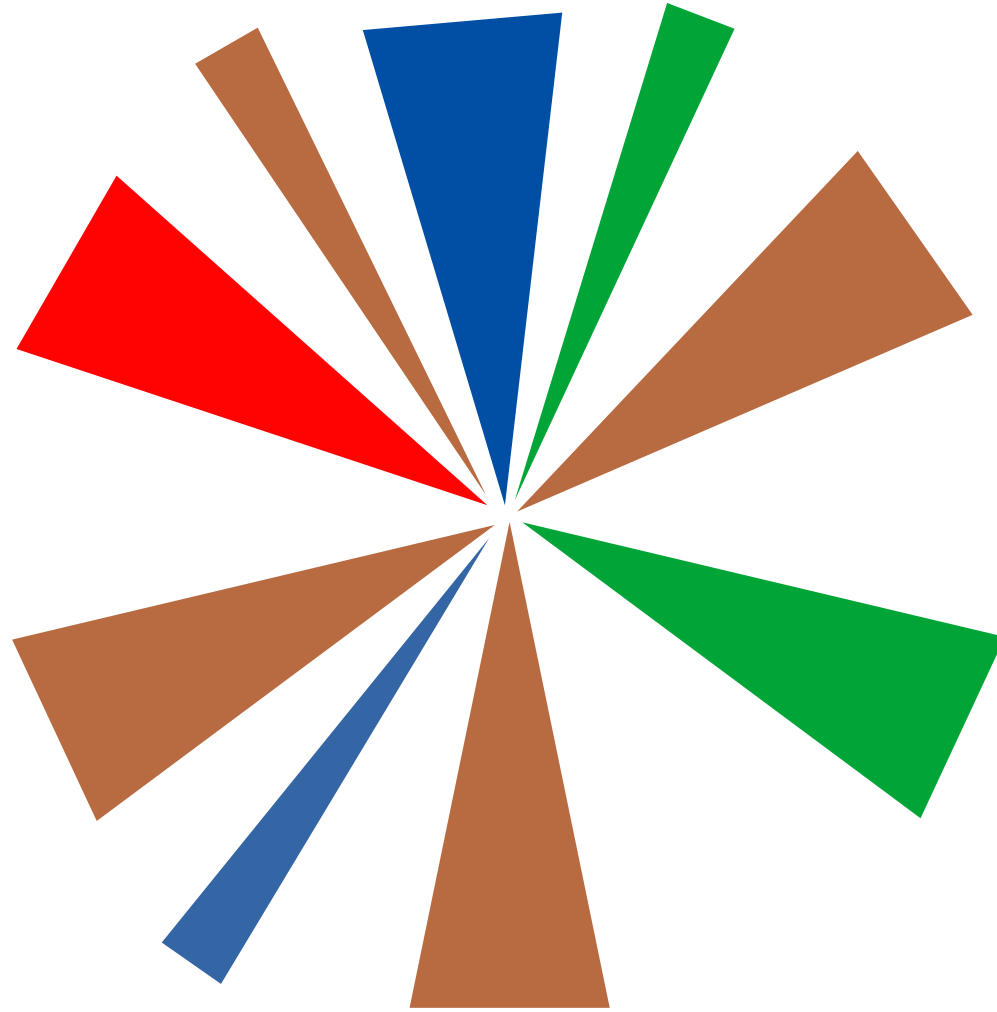








LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

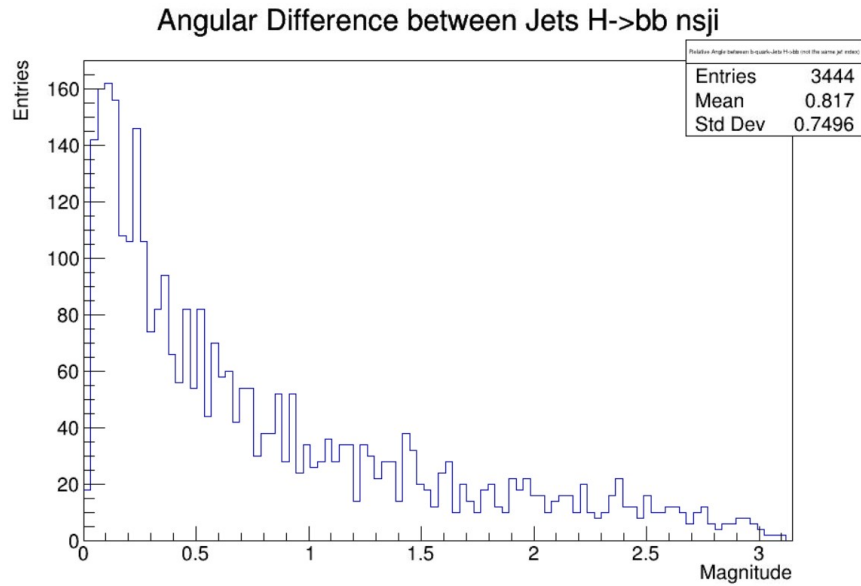




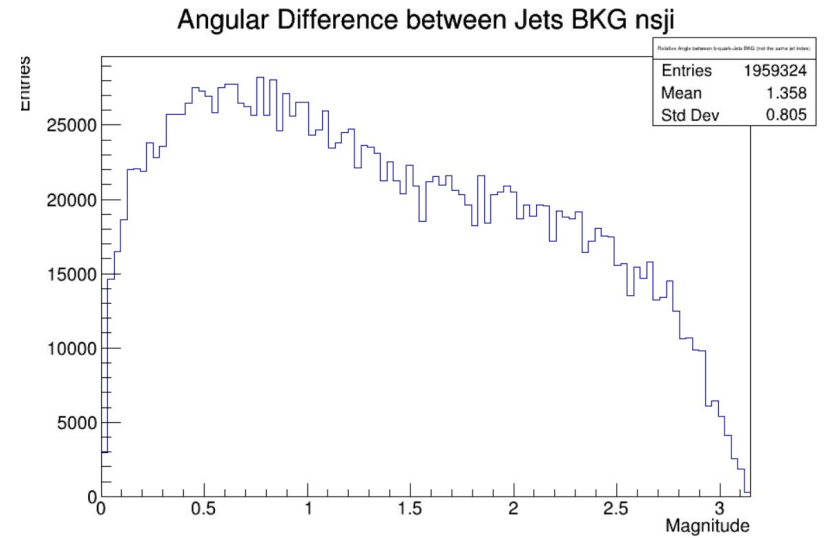
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Preliminary Analysis

Angle between
Jets



Delta R between
Jets



Relative
Mass

Total
JetMass

Preliminary Analysis

Other kinetic variables

Relative
Mass

Total
Jetmass

DeltaR

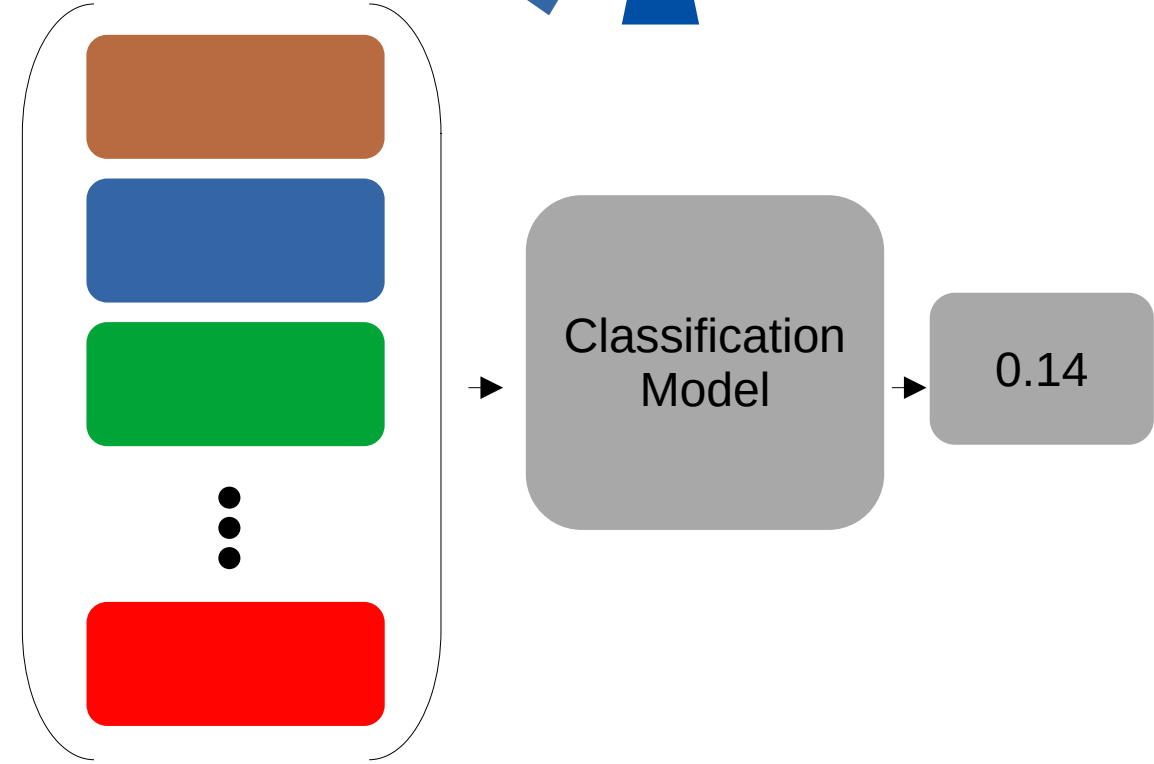
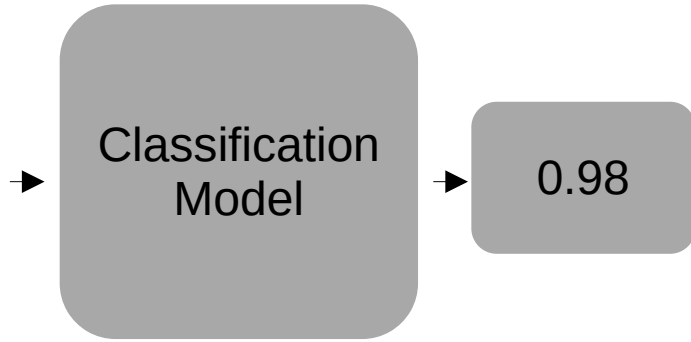
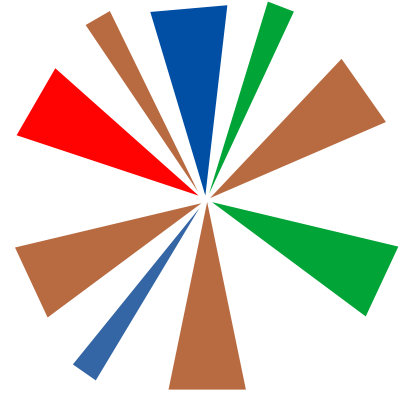
Px of Jet 1

Px of Jet 2

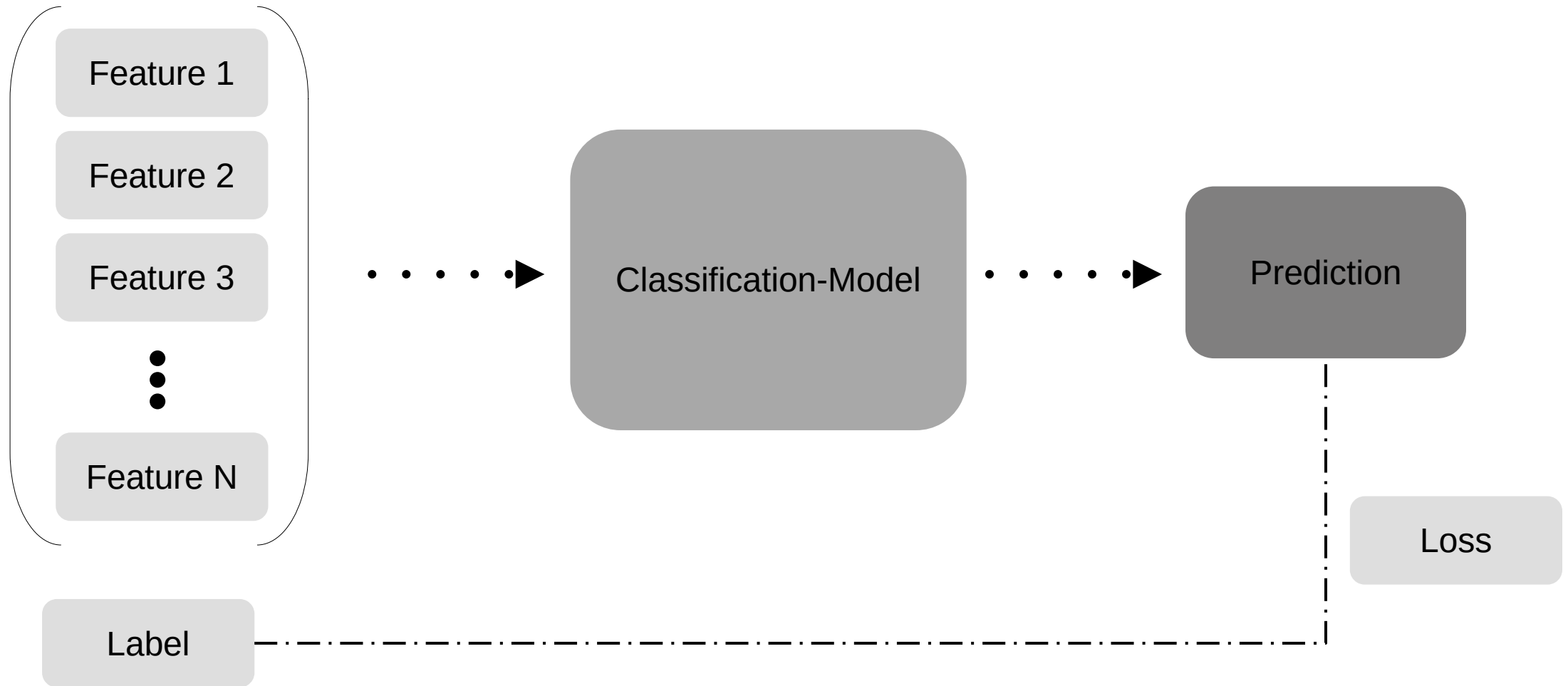
...

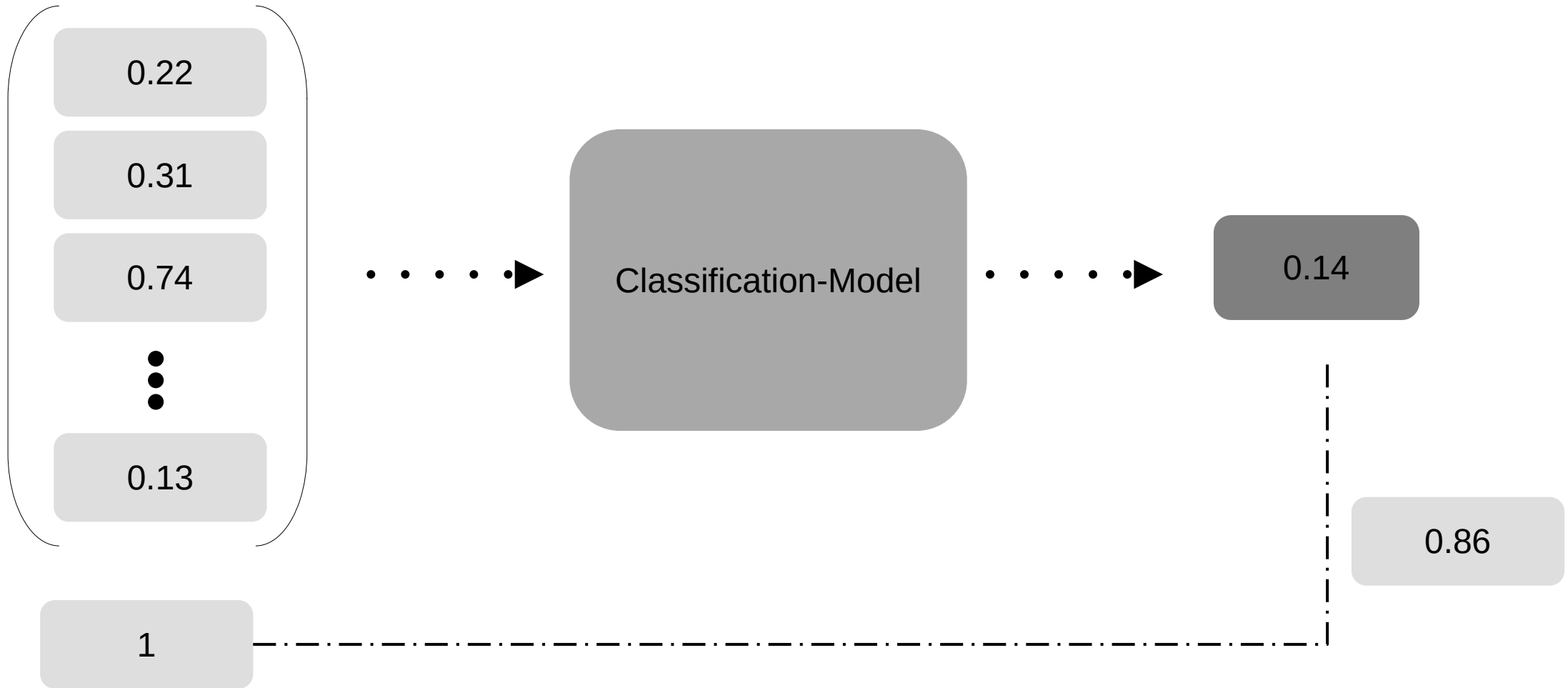
Label

Inputdata

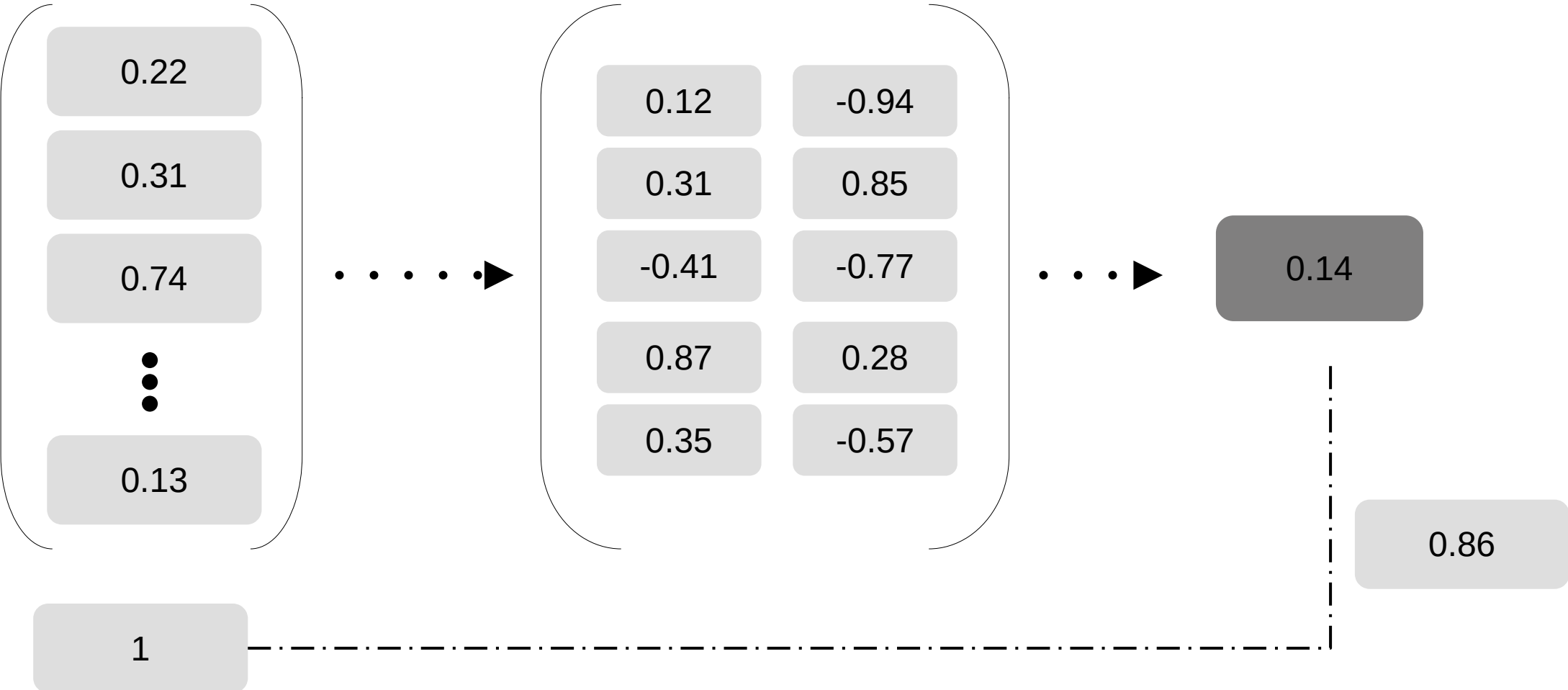


Intermission: Recap on Neural Network

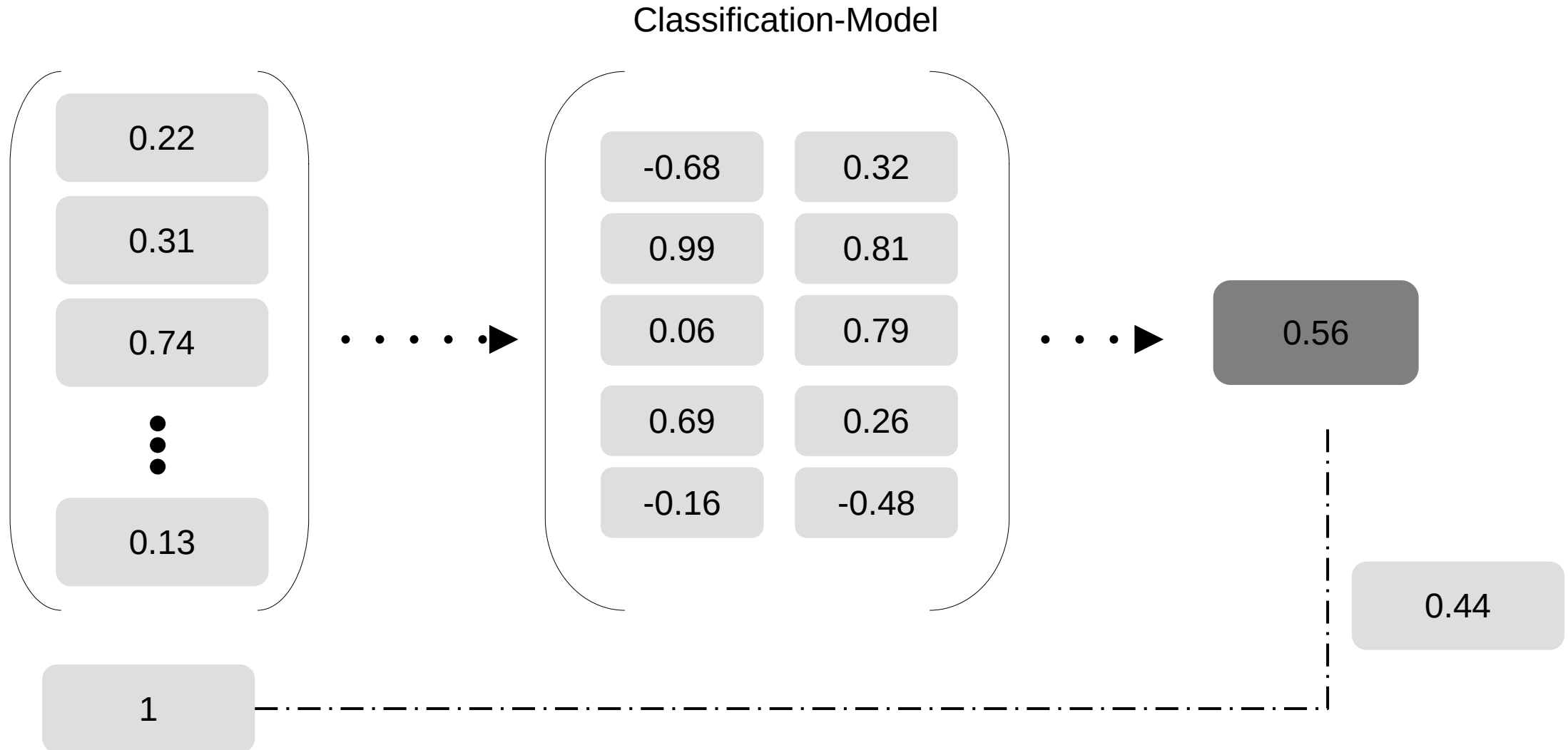




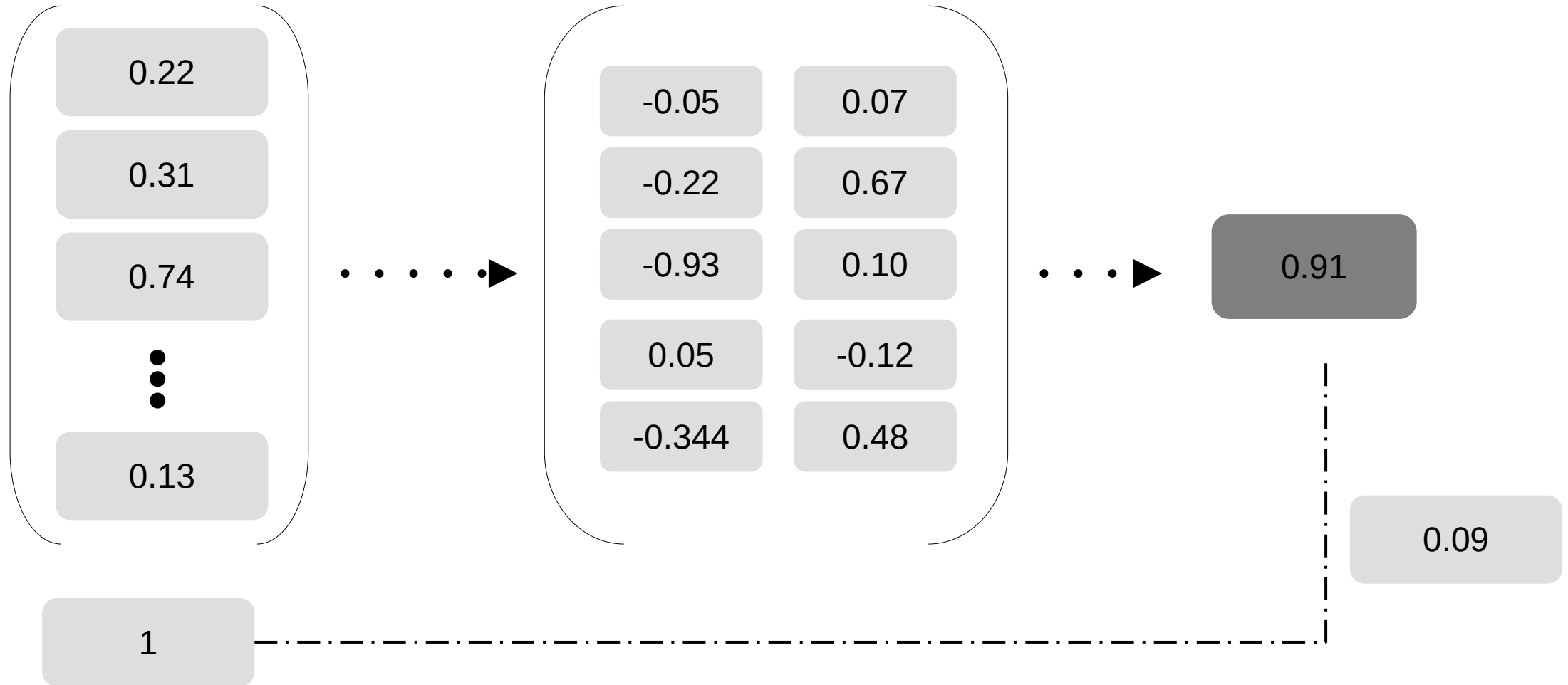
Classification-Model



Binary Classification

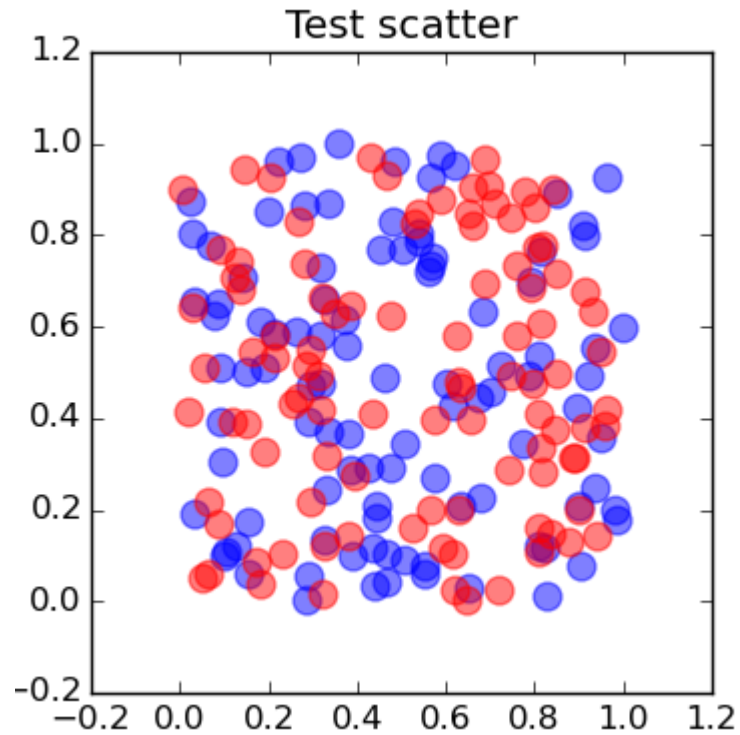


Classification-Model



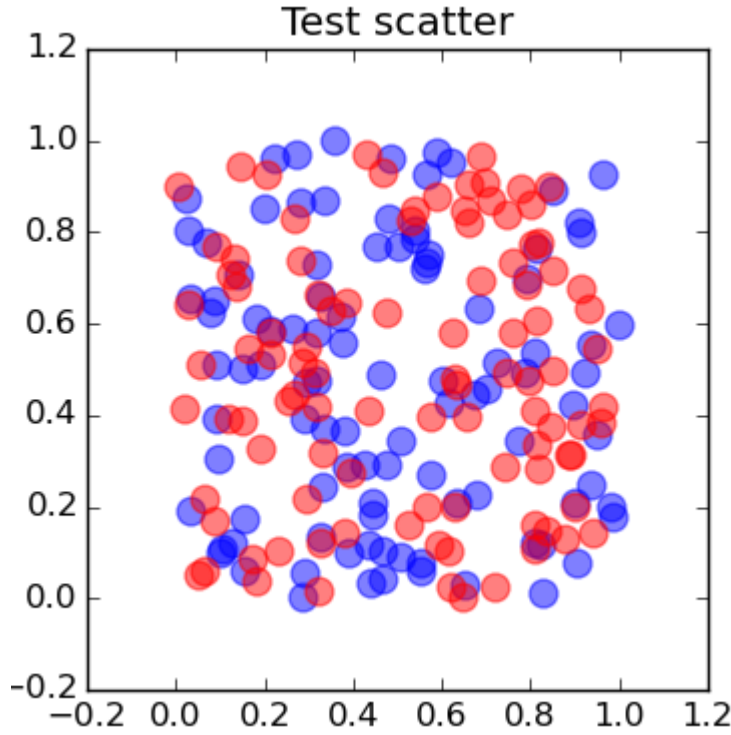
Triplett Learning

[TASK]
**Make a model that can tell the difference between these
red dots and these blue dots**

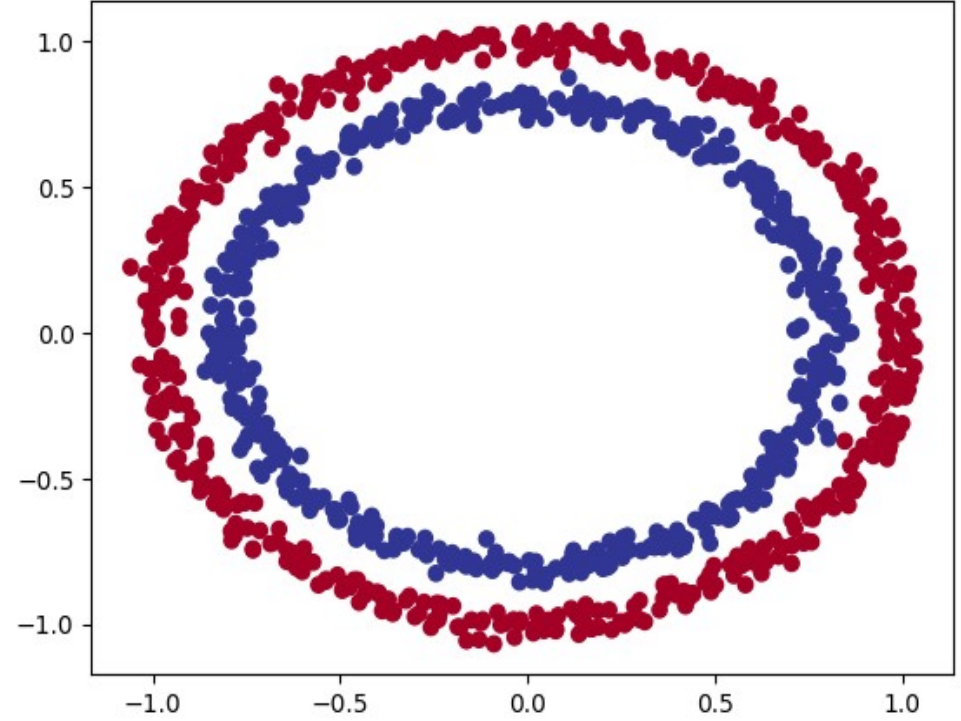
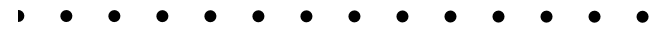


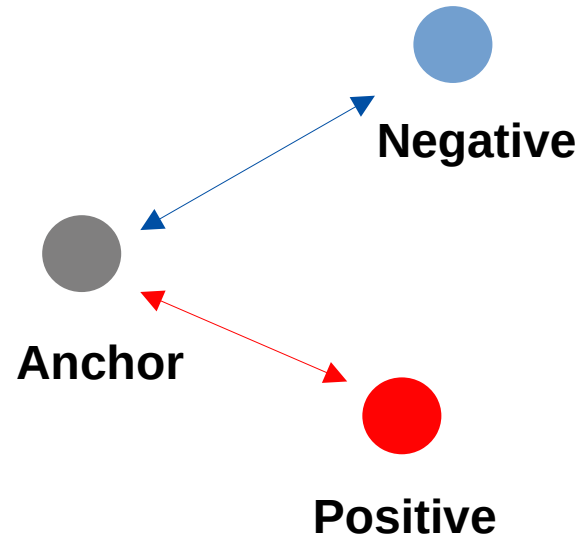
Triplet Learning-Intuition

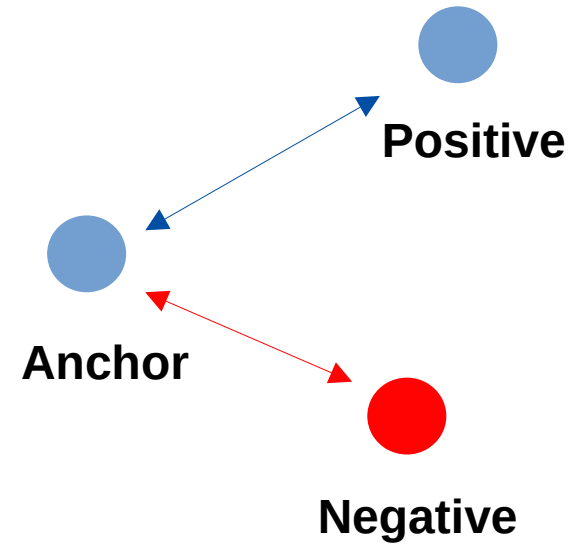
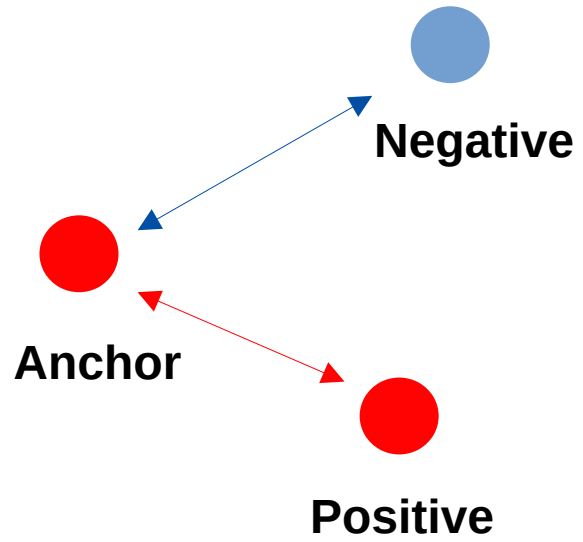
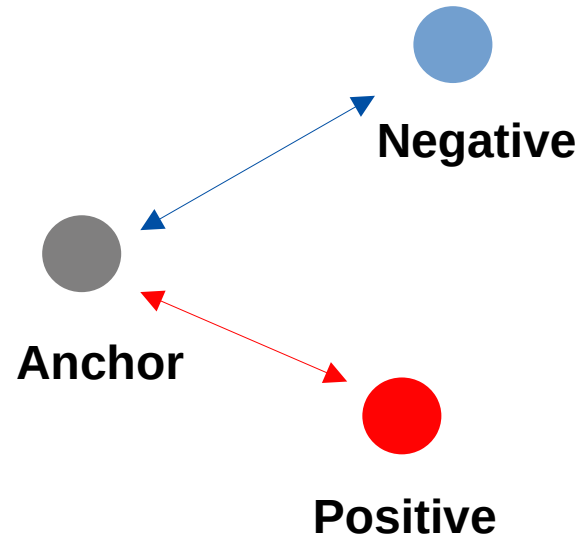
We want to find a coordinate transformation that groups datapoints of the same class together

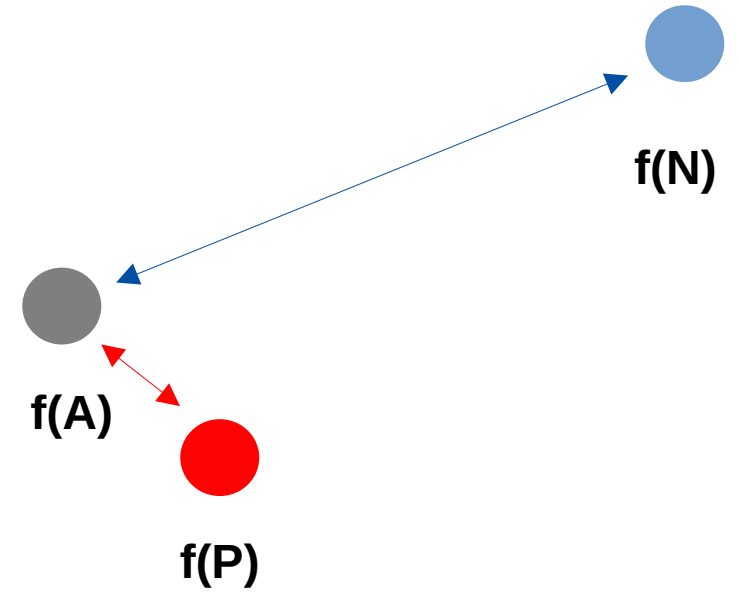
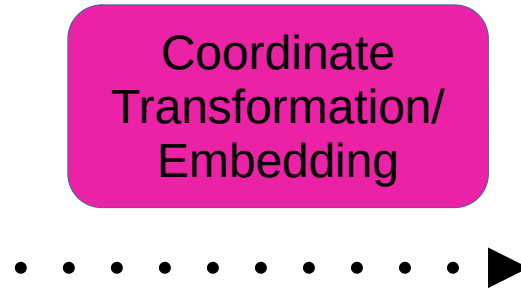
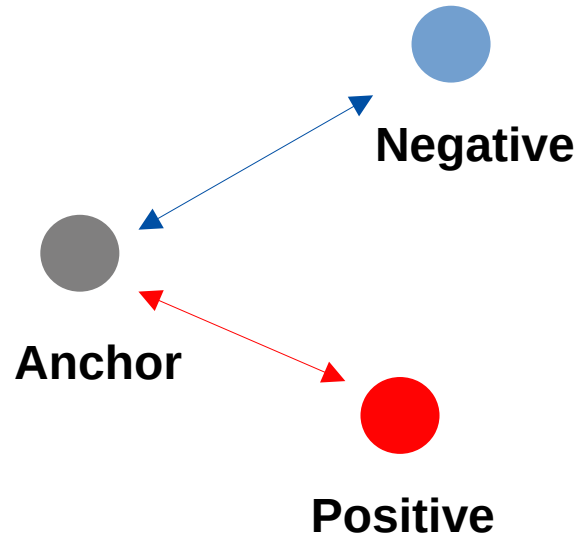


Coordinate Transformation/ Embedding

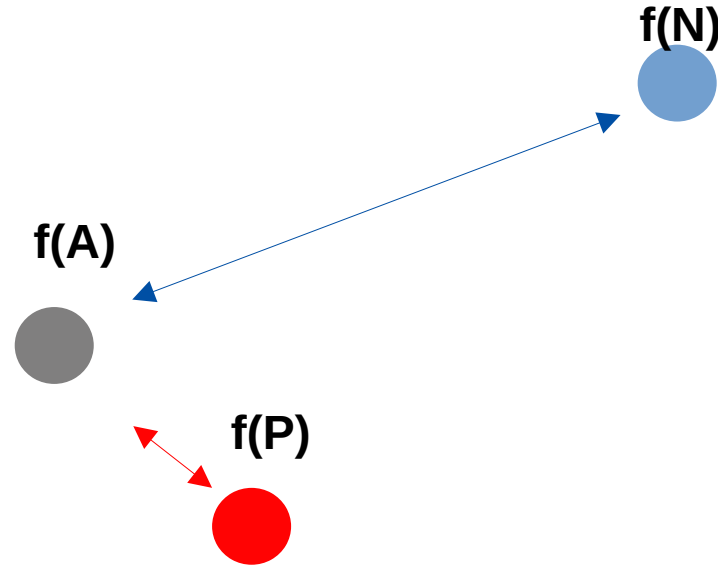








Triplet Loss Function



$L =$

$D(f(A), f(P))$

-

$D(f(A), f(N))$

The **Triplet Loss function / Contrastive Loss** is at a minimum with a large $D(f(A), f(N))$ and a small $D(f(A), f(P))$

$$L = D(f(A), f(P)) - D(f(A), f(N)) \sim 0$$

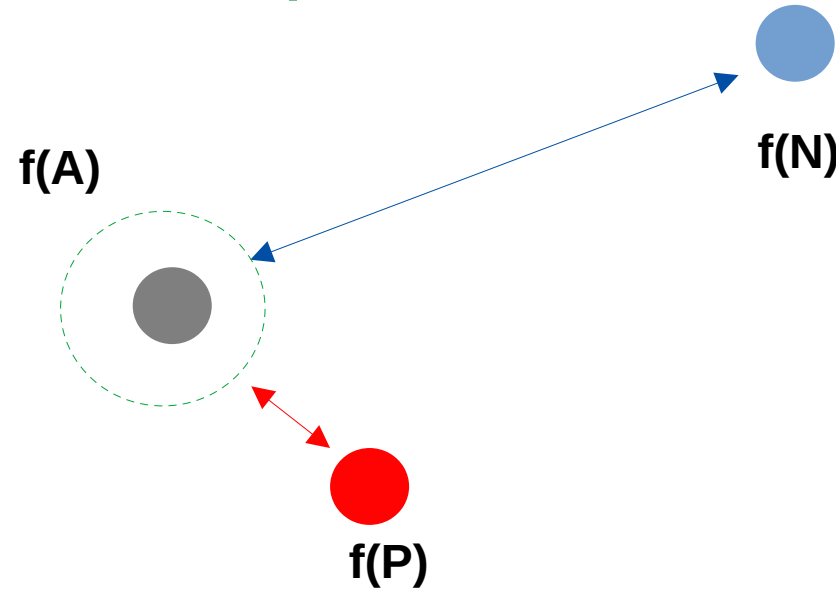
for

$$f(A) \sim f(P) \sim f(N)$$



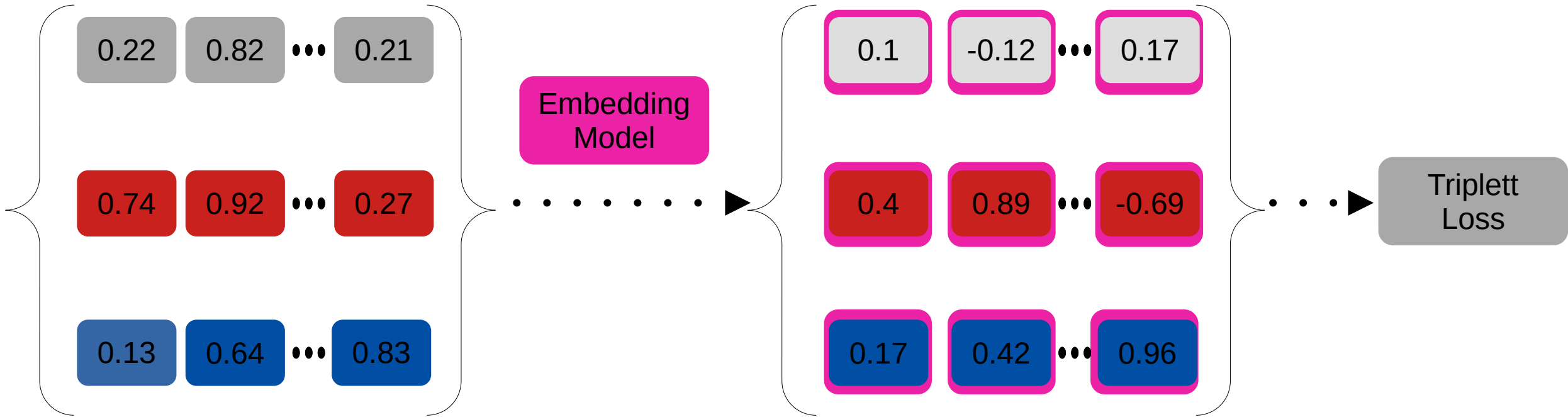
Not ideal for categorization

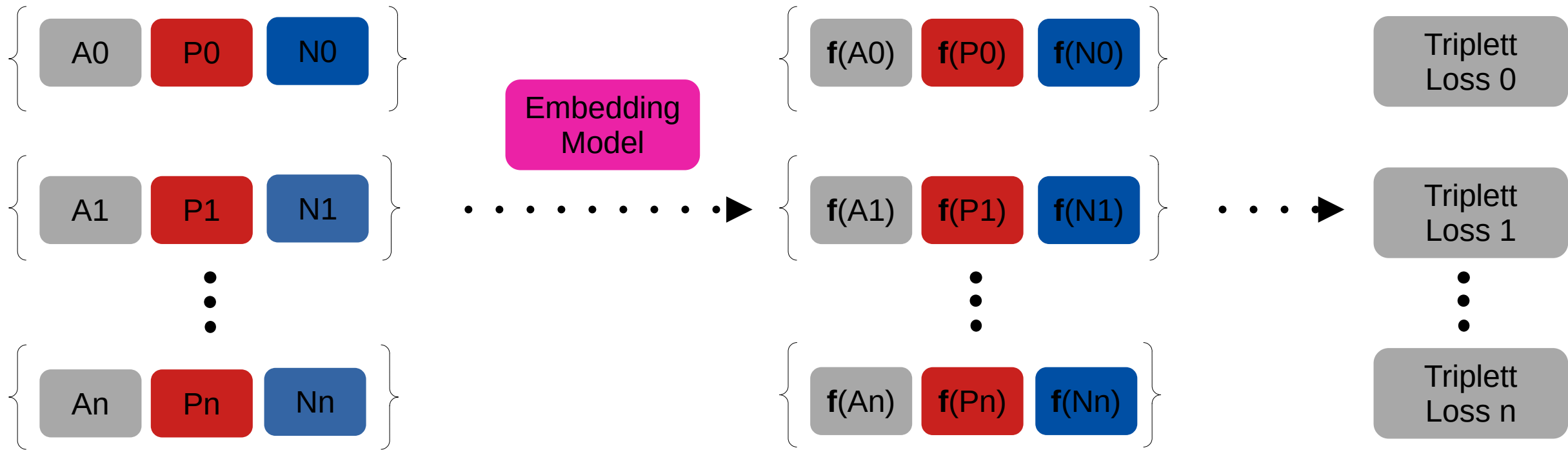
Triplet Loss Function

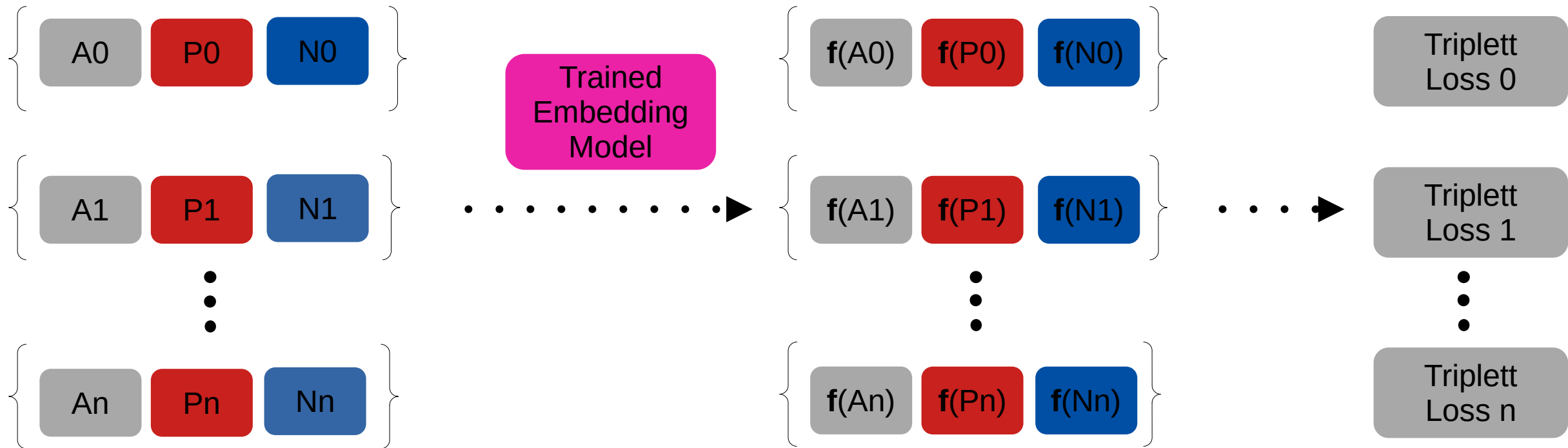


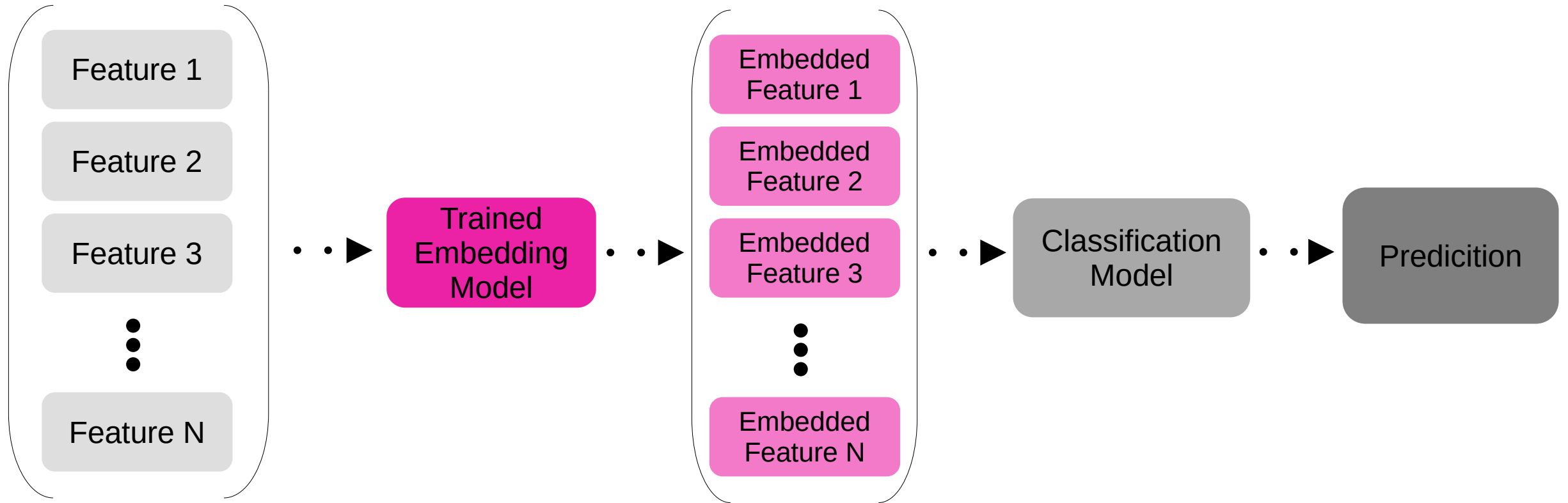
$$L = \max(D(f(A), f(P)) - D(f(A), f(N)) + m , 0)$$

The **margin m** is a hyperparameter that prevents collapse and may reduce expressiveness

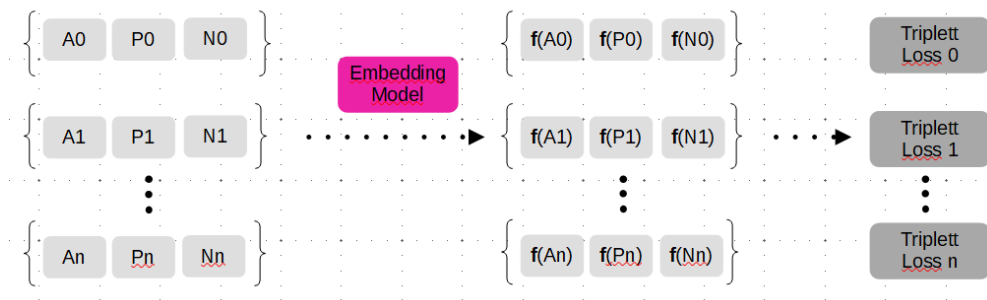






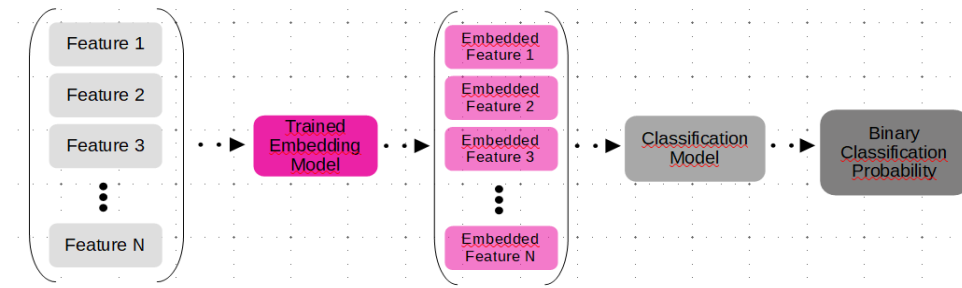


Where I currently am



Current State

Setup the code triplett creation and triplett training pipeline



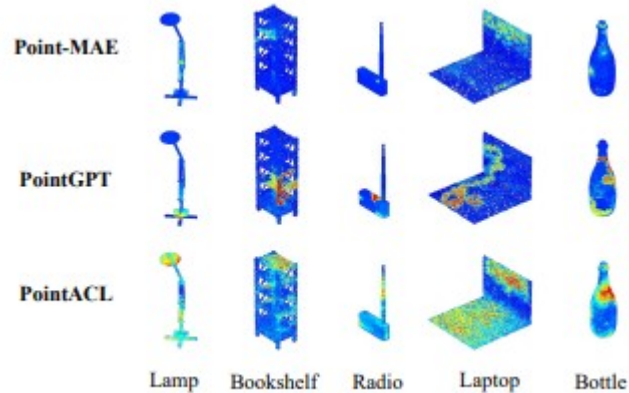
Setup the code to embed the inputfeatures and train a classification model on it

**Preliminary Analysis is almost done
Need to prepare a bigger BKG and SIG dataset**

What do I want to show

| | | | | | |
|-------------------|------------|--------------------|--------------------|--------------------|--------------------|
| Point-BERT [57] | CVPR 22 | 87.4 | 88.1 | 83.1 | 92.7 |
| +PointACL | - | 89.5 (+2.1) | 88.6 (+0.5) | 84.5 (+1.4) | 93.1 (+0.4) |
| Point-MAE [32] | ECCV 22 | 90.0 | 88.3 | 85.2 | 93.2 |
| +PointACL | - | 90.9 (+0.9) | 88.8 (+0.5) | 85.4 (+0.2) | 93.7 (+0.5) |
| PointGPT-S [5] | NeurIPS 23 | 91.6 | 90.0 | 86.9 | 93.3 |
| +PointACL | - | 92.3 (+0.7) | 91.6 (+1.6) | 87.1 (+0.2) | 93.5 (+0.2) |
| Point-MAE* [32] | ECCV 22 | 92.8 | 91.2 | 89.0 | 93.8 |
| +PointACL* | - | 93.1 (+0.3) | 91.7 (+0.5) | 89.2 (+0.2) | 94.1 (+0.3) |
| PointGPT-S* [5] | NeurIPS 23 | 93.4 | 92.4 | 89.2 | 94.0 |
| +PointACL* | - | 94.5 (+1.1) | 93.5 (+1.1) | 89.9 (+0.7) | 94.1 (+0.1) |

Performs better or at least competitively against non-embedded data



Is more resistant against noisy data

Further investigation

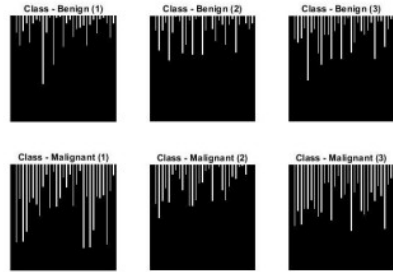


Figure 3: Bar graph for some data examples of Breast Cancer dataset.

Anuraganand Sharma (2020) Nonimage Data Classification with Convolutional Neural Networks
-<https://arxiv.org/pdf/2007.03218v1>

„We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done.“

Rich Sutton (2019) The Bitter Lesson
-<http://www.incompleteideas.net/InIdeas/BitterLesson.html>

Working with 2-D data formed from tabular data

Working with more „raw“ data according to the „Bitter Truth“



Resources

-The Nobel Prize in Physics 2013. NobelPrize.org. Nobel Prize Outreach AB 2025. Sat. 4 Jan 2025. <<https://www.nobelprize.org/prizes/physics/2013/summary/>>

- <https://home.cern/science/physics/higgs-boson/how>

-Dimensionality Reduction by Learning an Invariant Mapping

Raia Hadsell, Sumit Chopra, Yann LeCun

The Courant Institute of Mathematical Sciences



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Bao Tai, Le
Luise-Kieselbachplatz 31 · 81377 München ·
Tai.Le@physik.lmu.de



Why use Triplet-Training?

Show the new neural network graph
Look at certain paper for inspiration