

MEAN TEACHERS ARE BETTER ROLE MODELS

Weight-averaged consistency targets improve
semi-supervised deep learning results

ANTTI TARVAINEN & HARRI VALPOLA



RESULTS

CIFAR-10 WITH 4000 LABELS

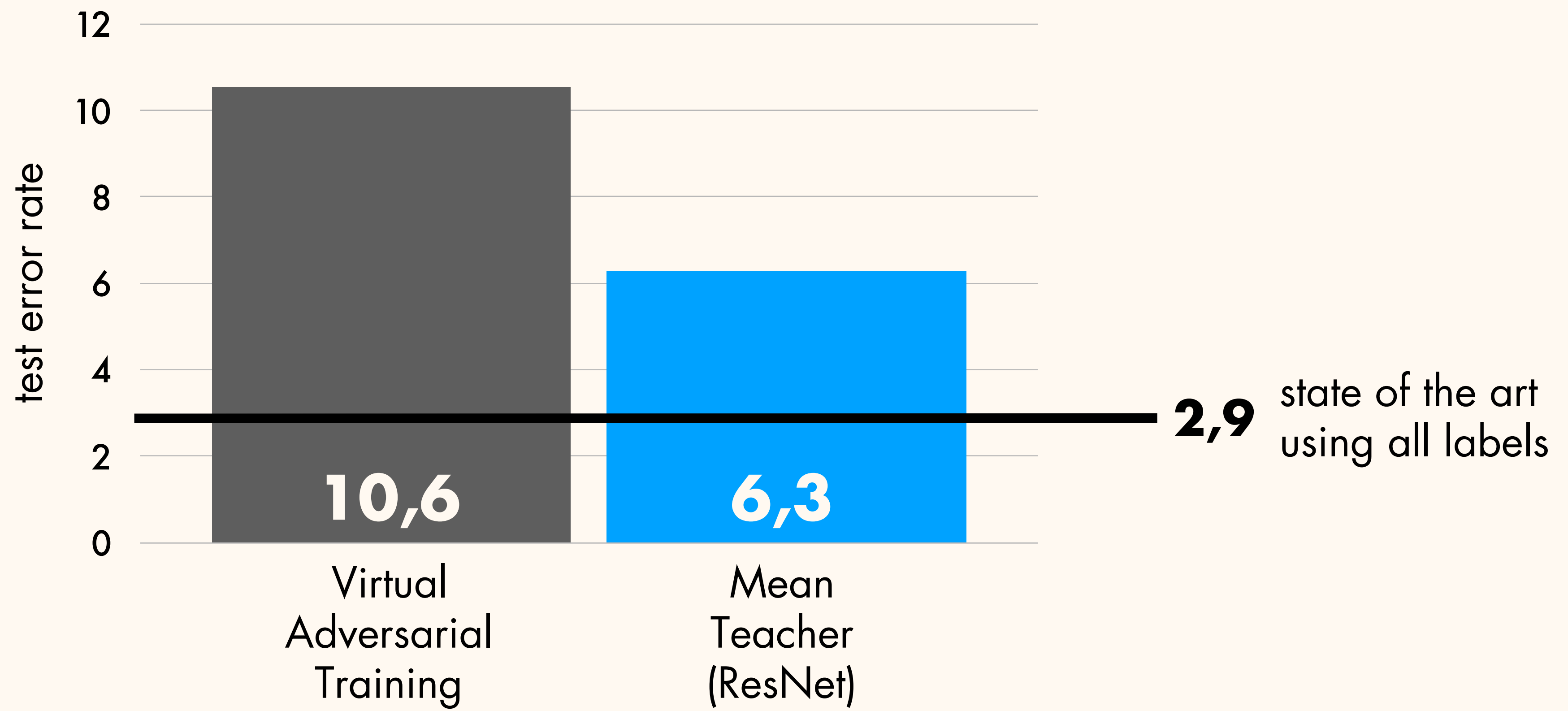


truck



**4000
labeled**

CIFAR-10 WITH 4000 LABELS



IMAGENET 2012 WITH 10% OF THE LABELS

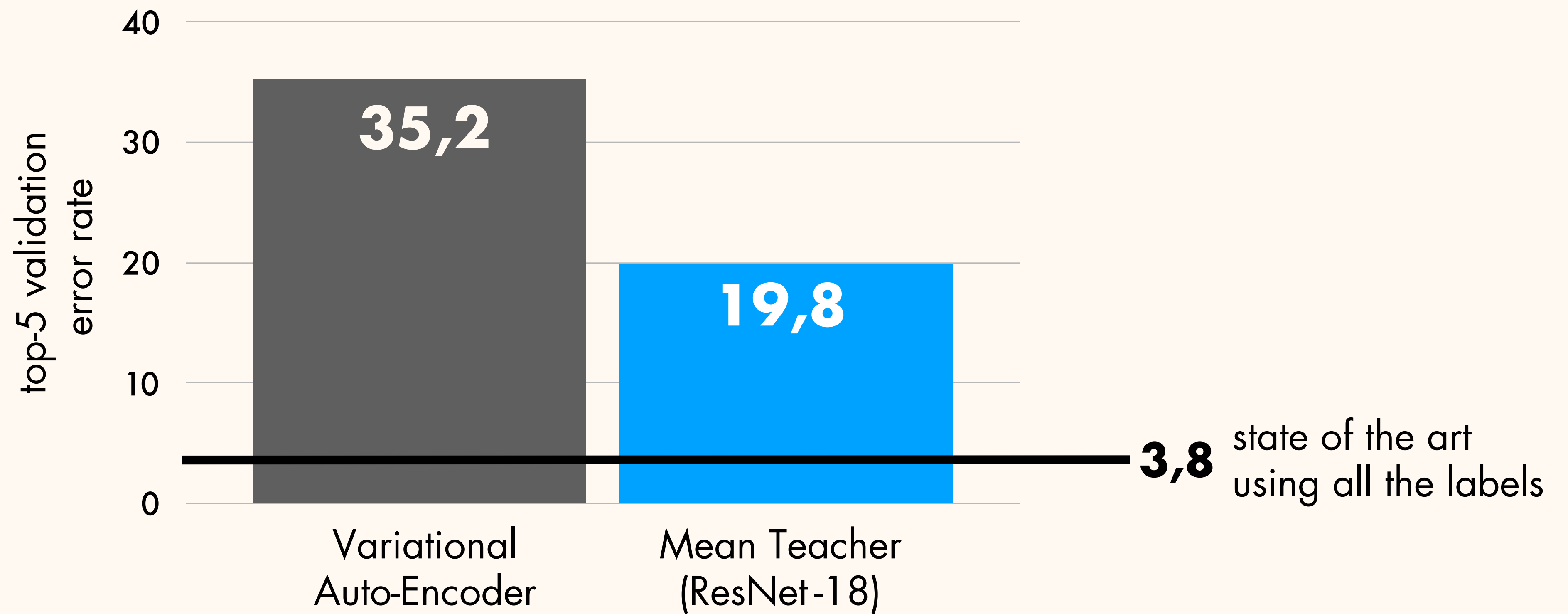


brambling

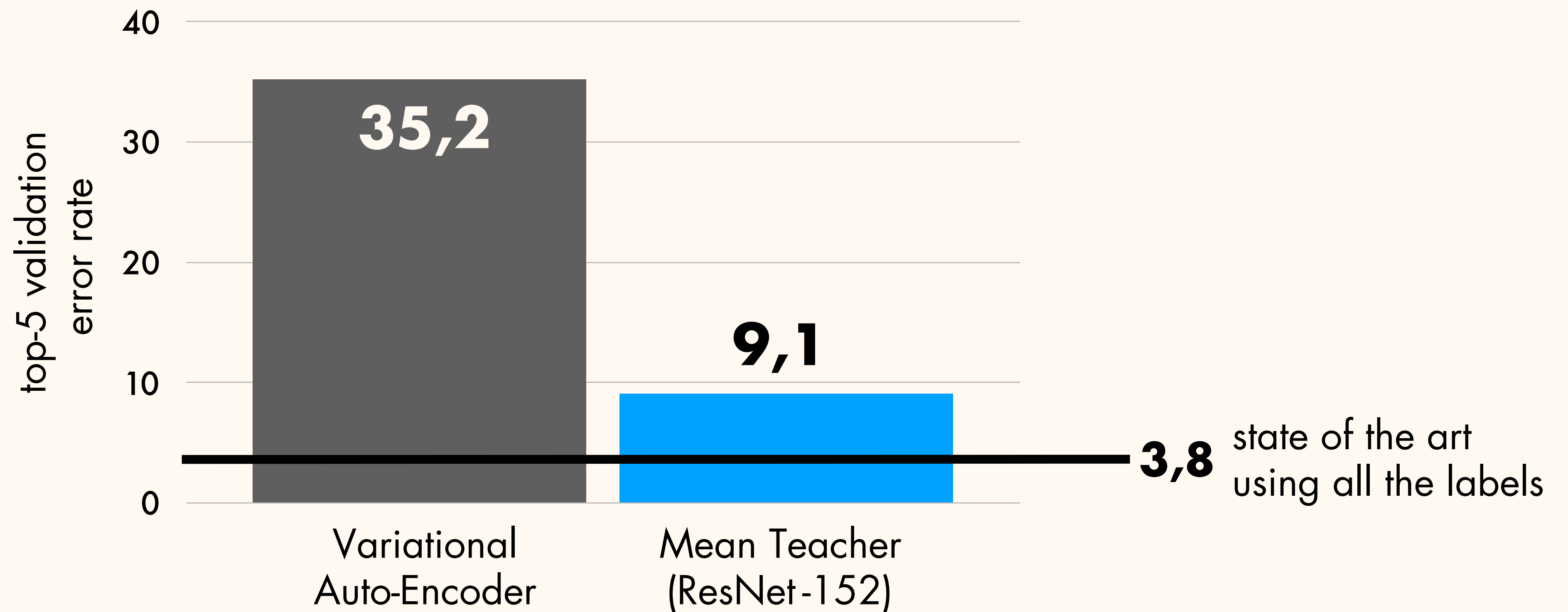
**1280000
images**

**10%
labeled**

IMAGENET 2012 WITH 10% OF THE LABELS



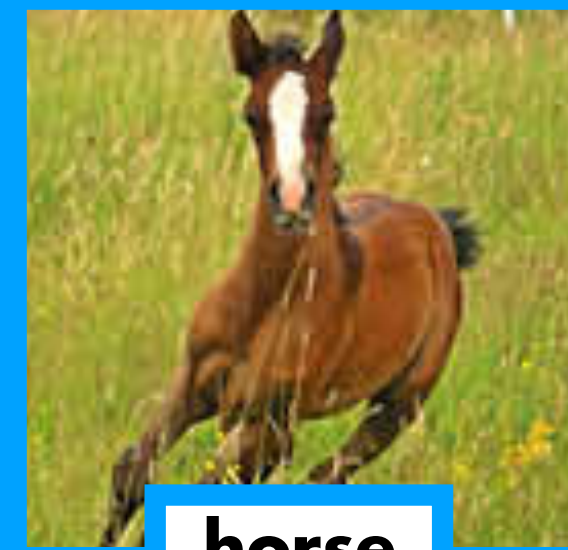
IMAGENET 2012 WITH 10% OF THE LABELS



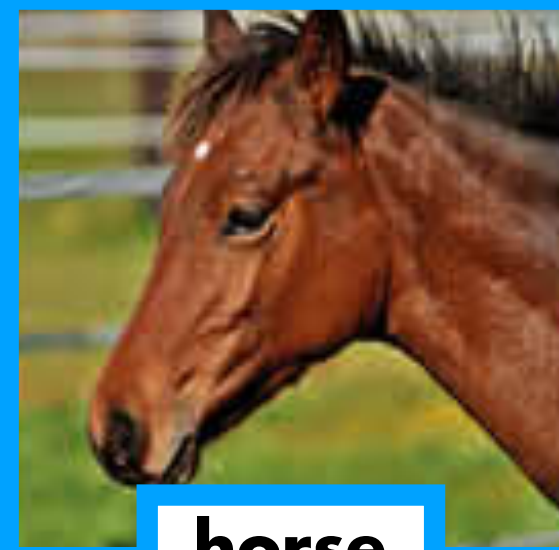
(these results included in the Arxiv version of the paper)

HOW THE MEAN TEACHER WORKS





horse



horse



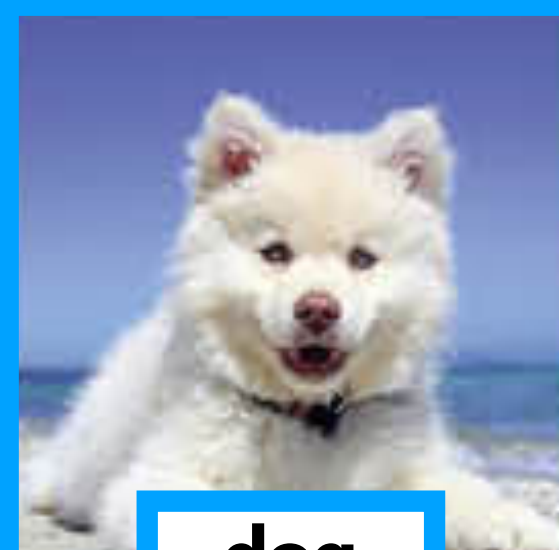
cat



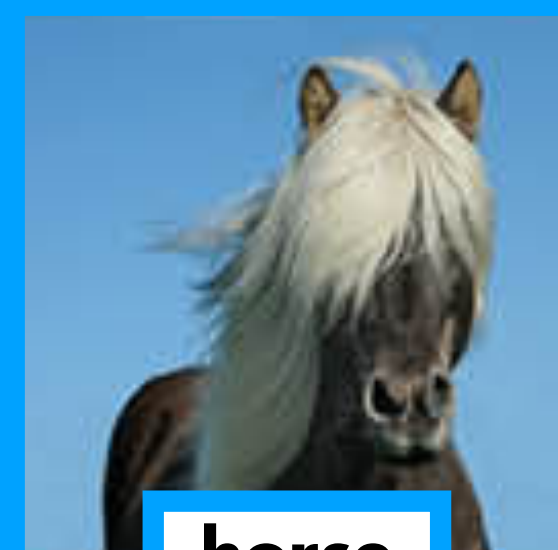
cat



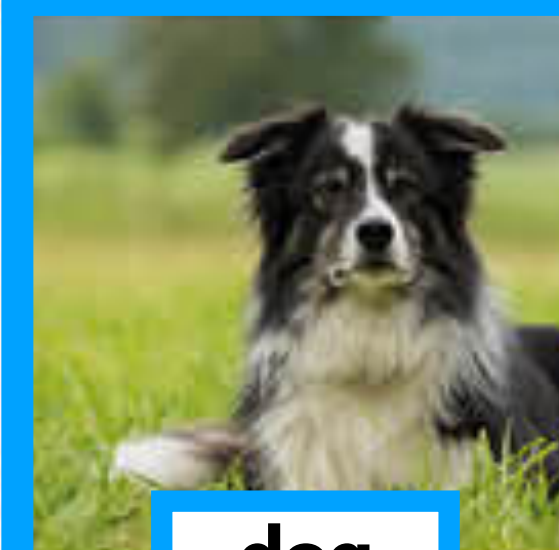
cat



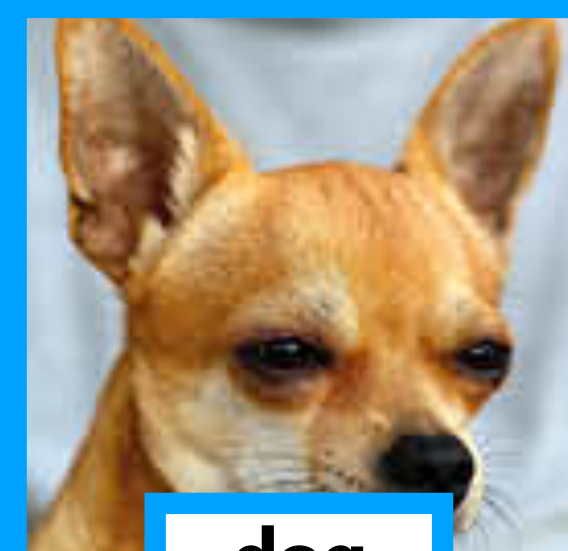
dog



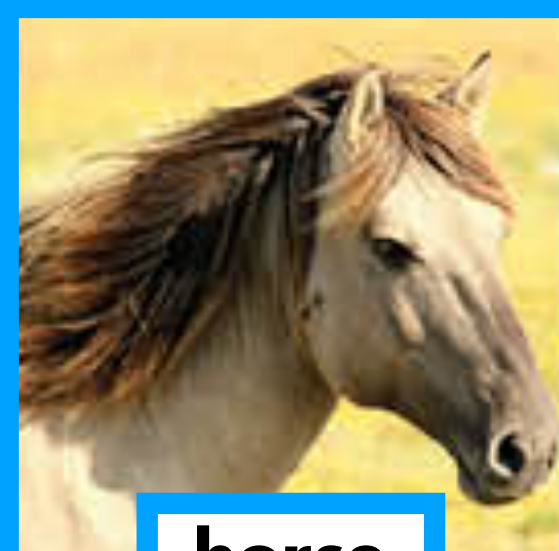
horse



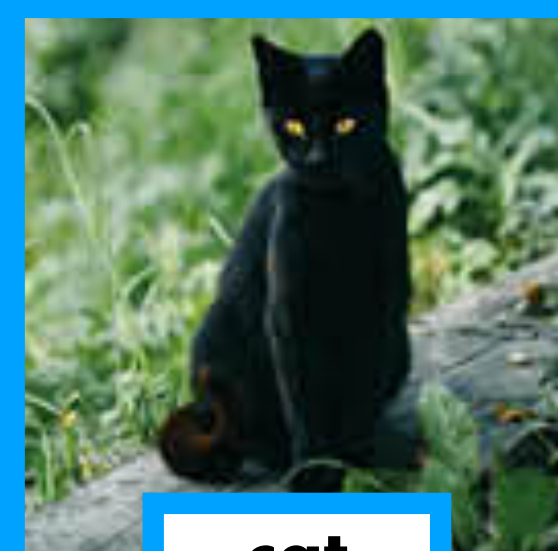
dog



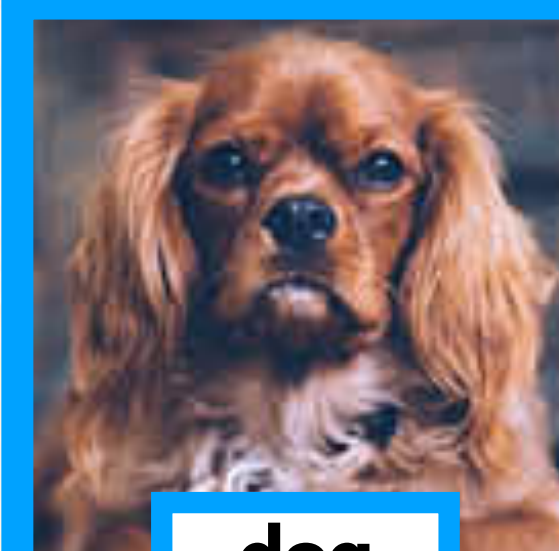
dog



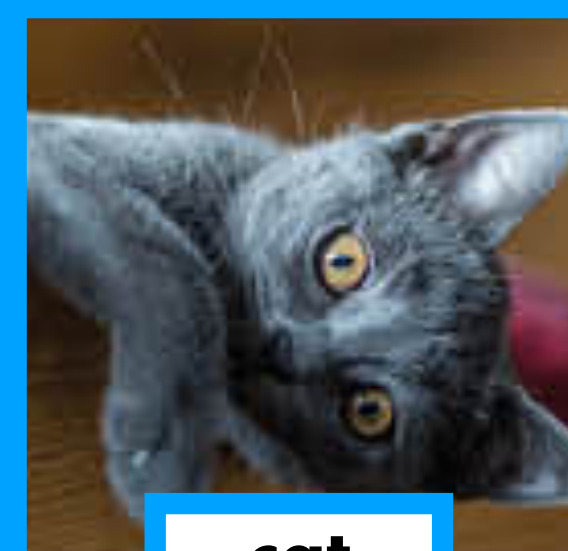
horse



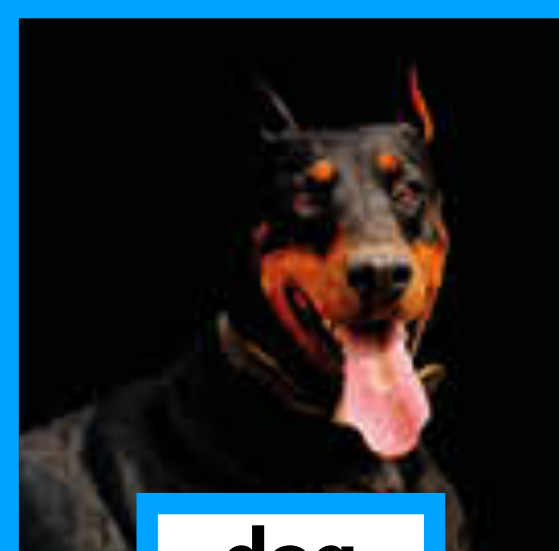
cat



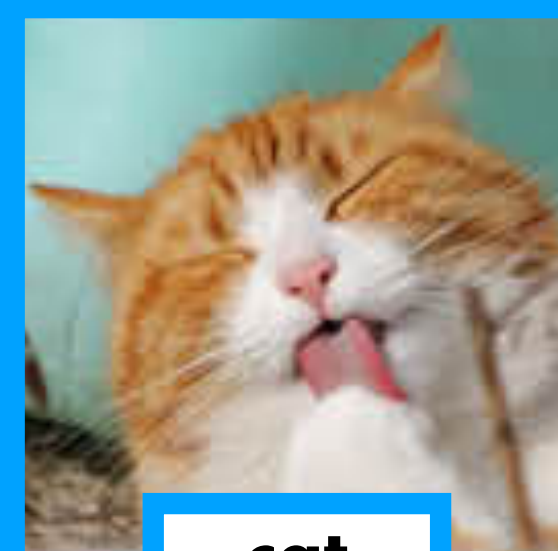
dog



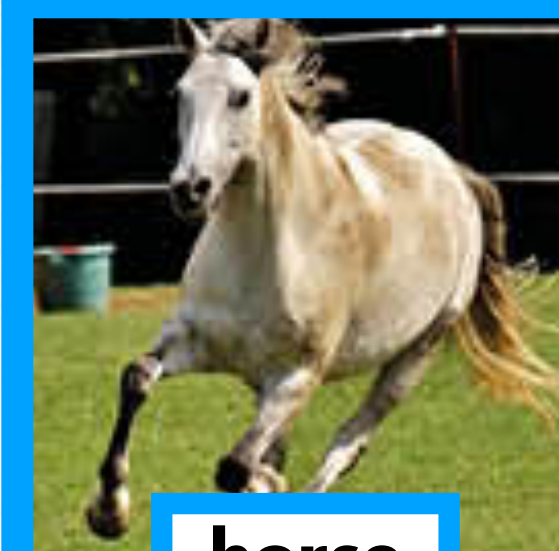
cat



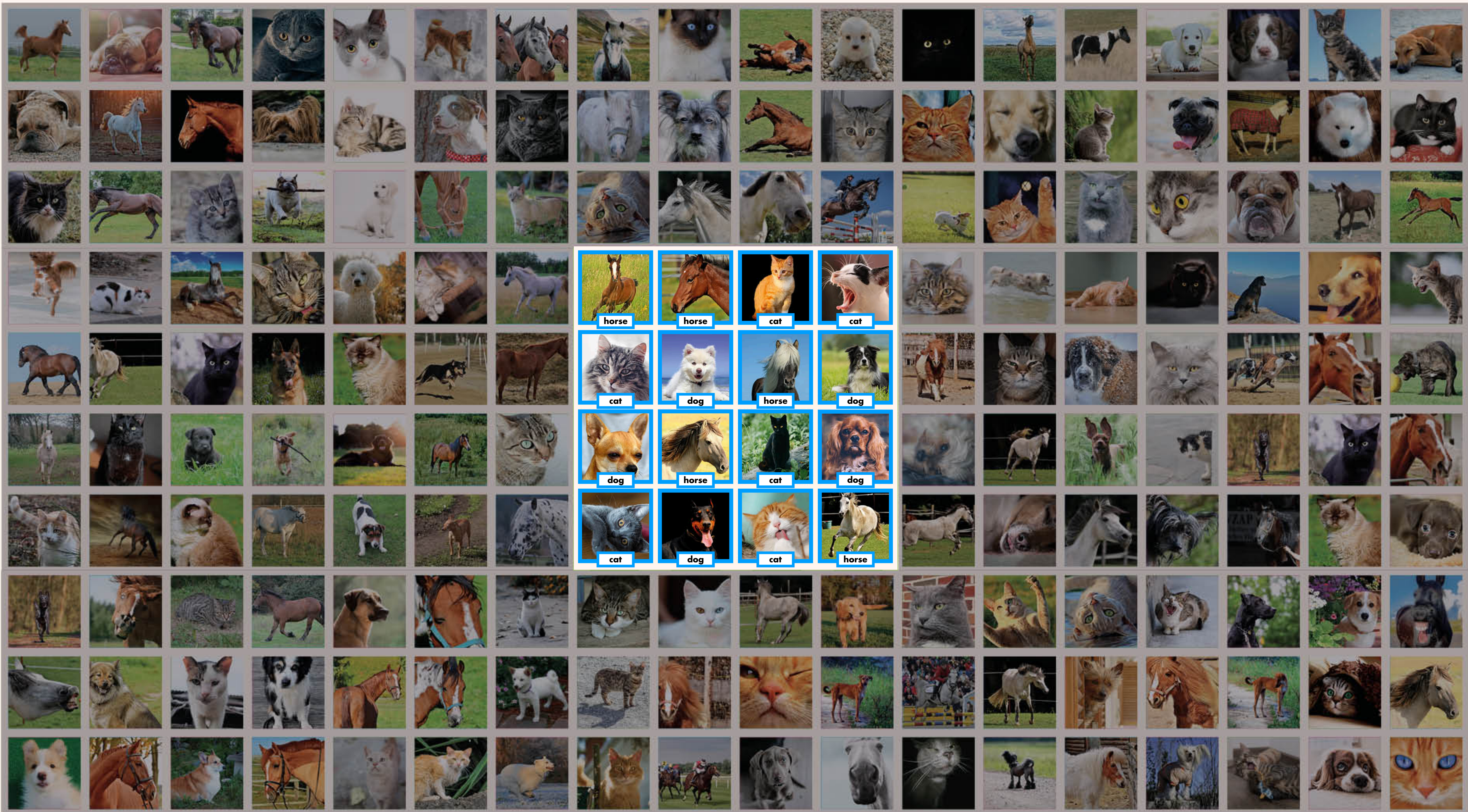
dog

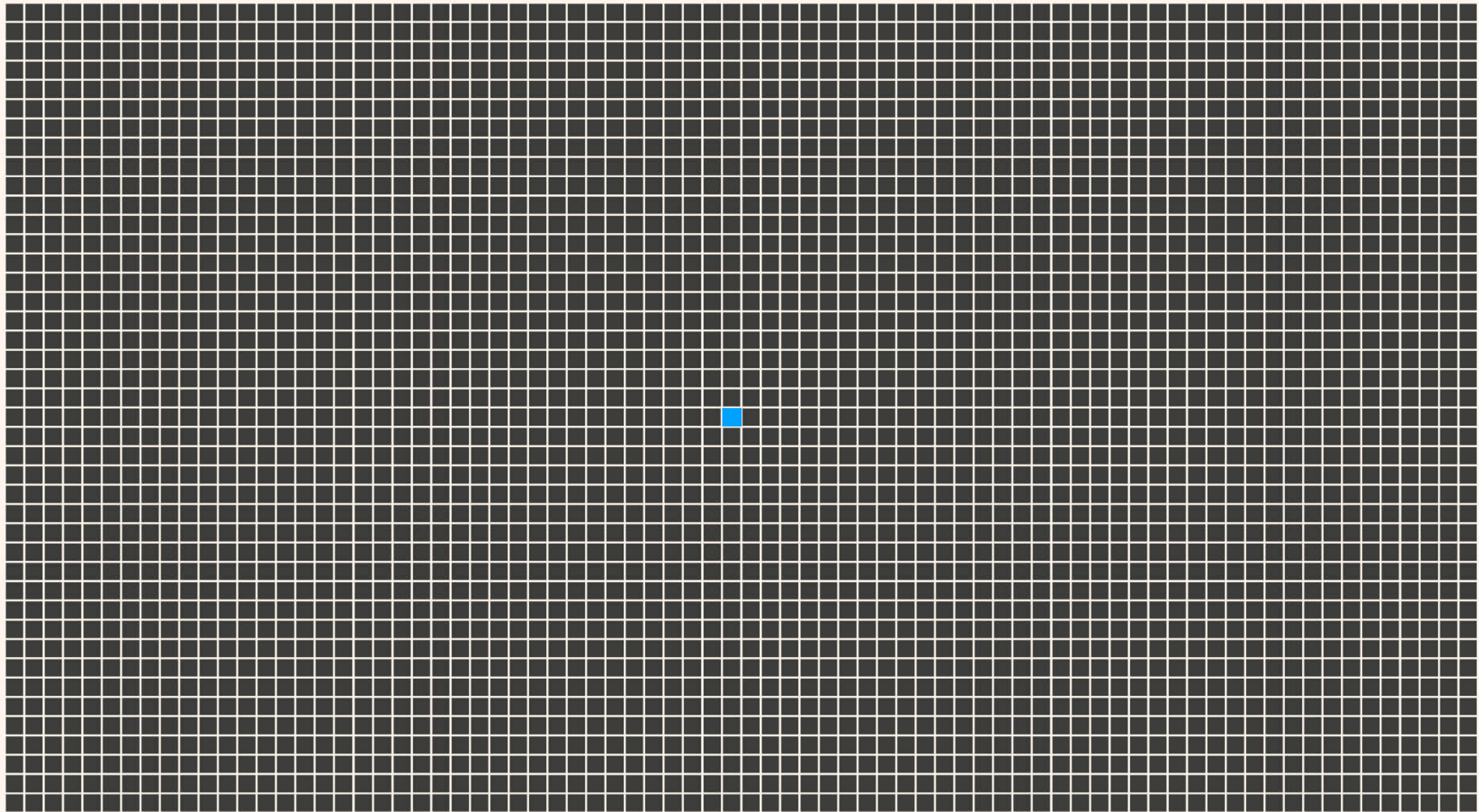


cat

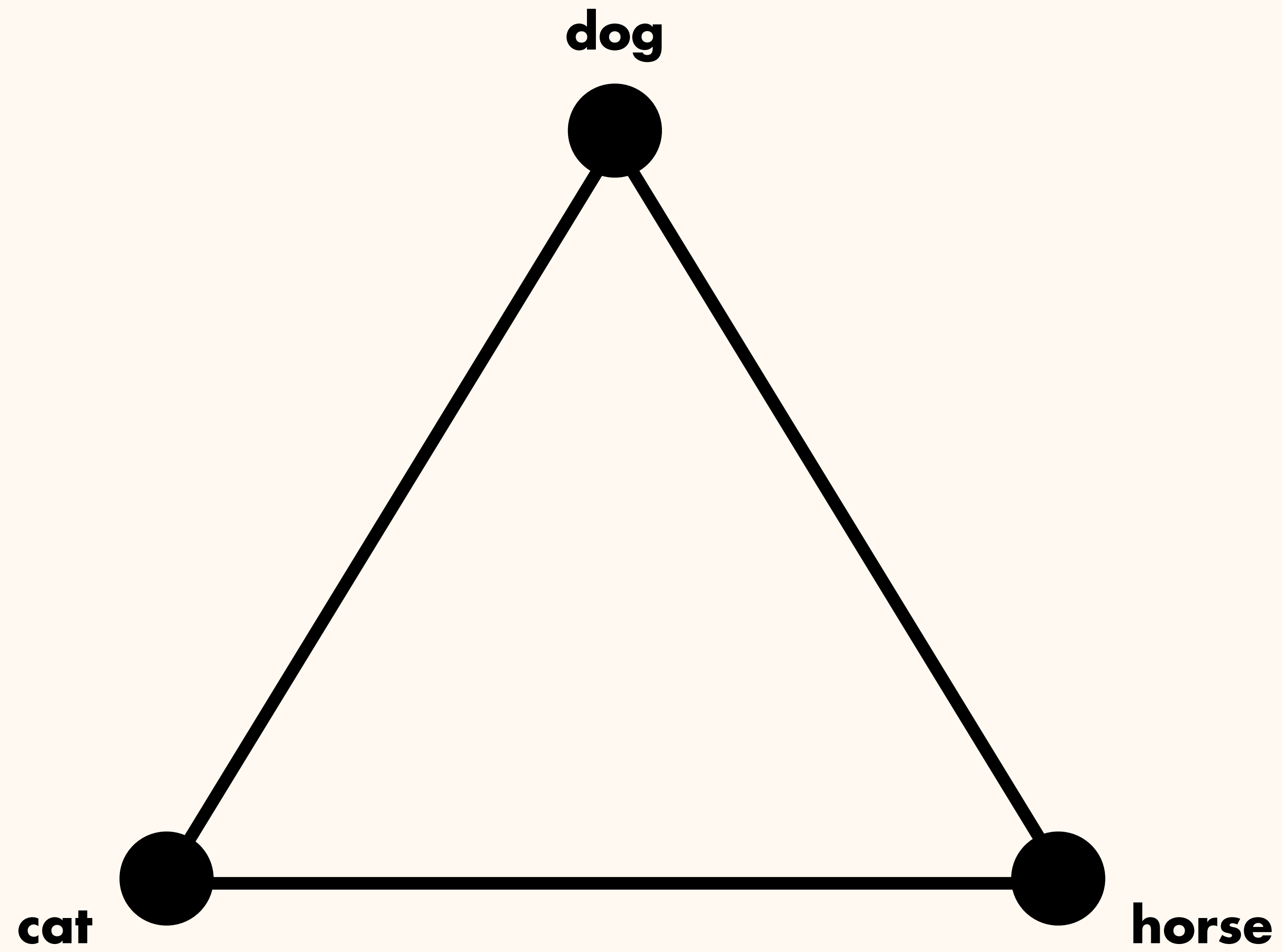


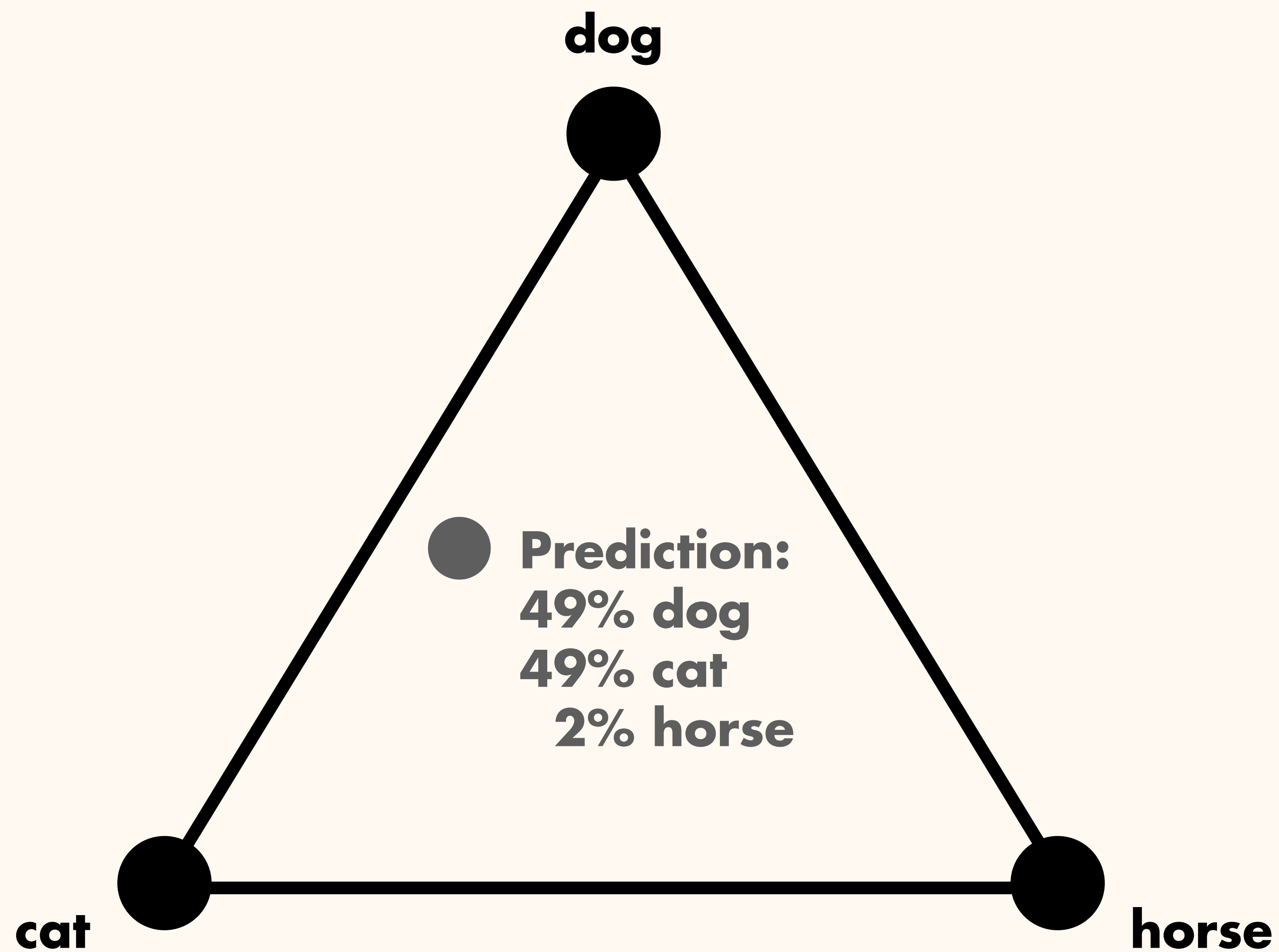
horse

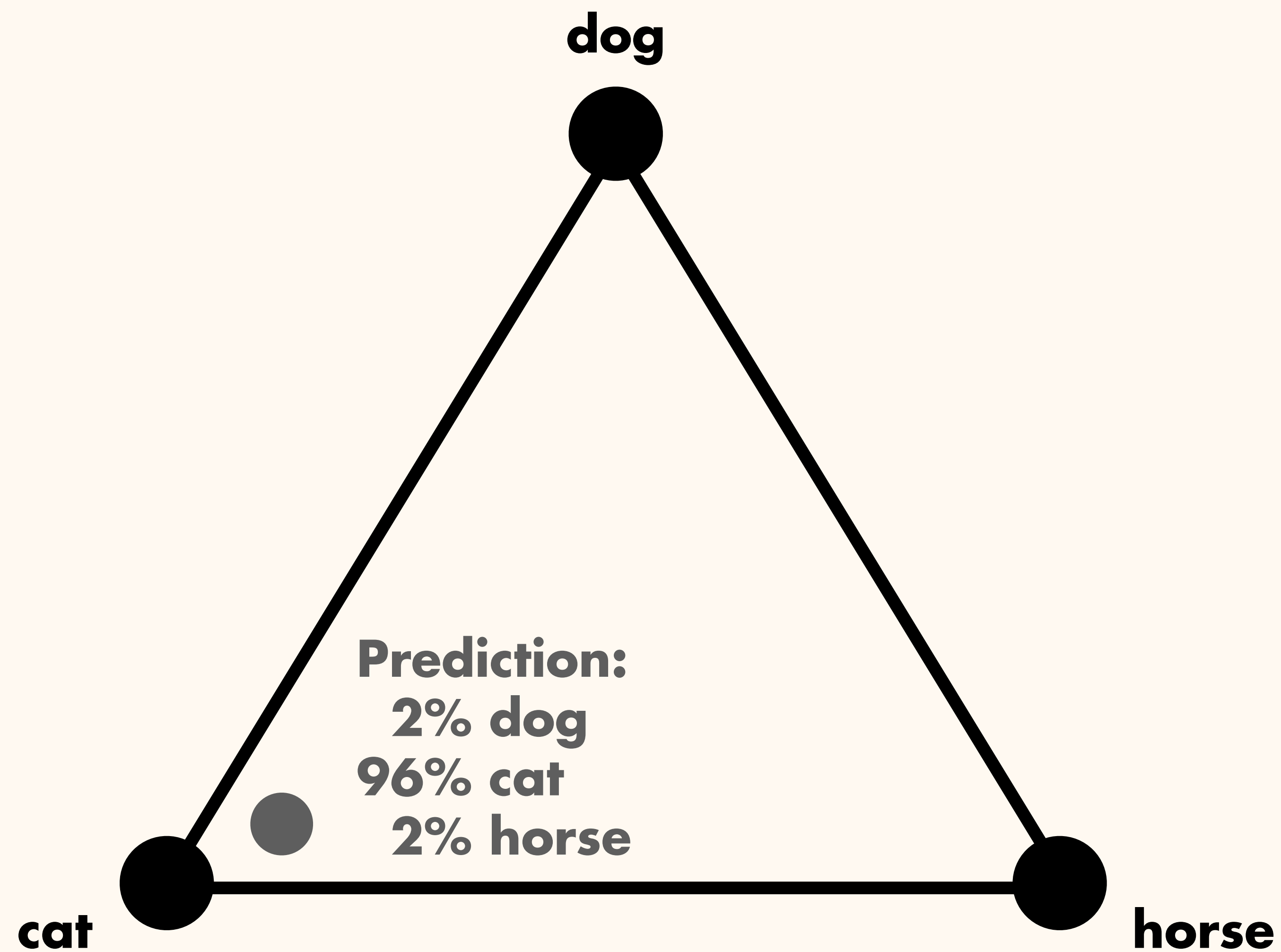


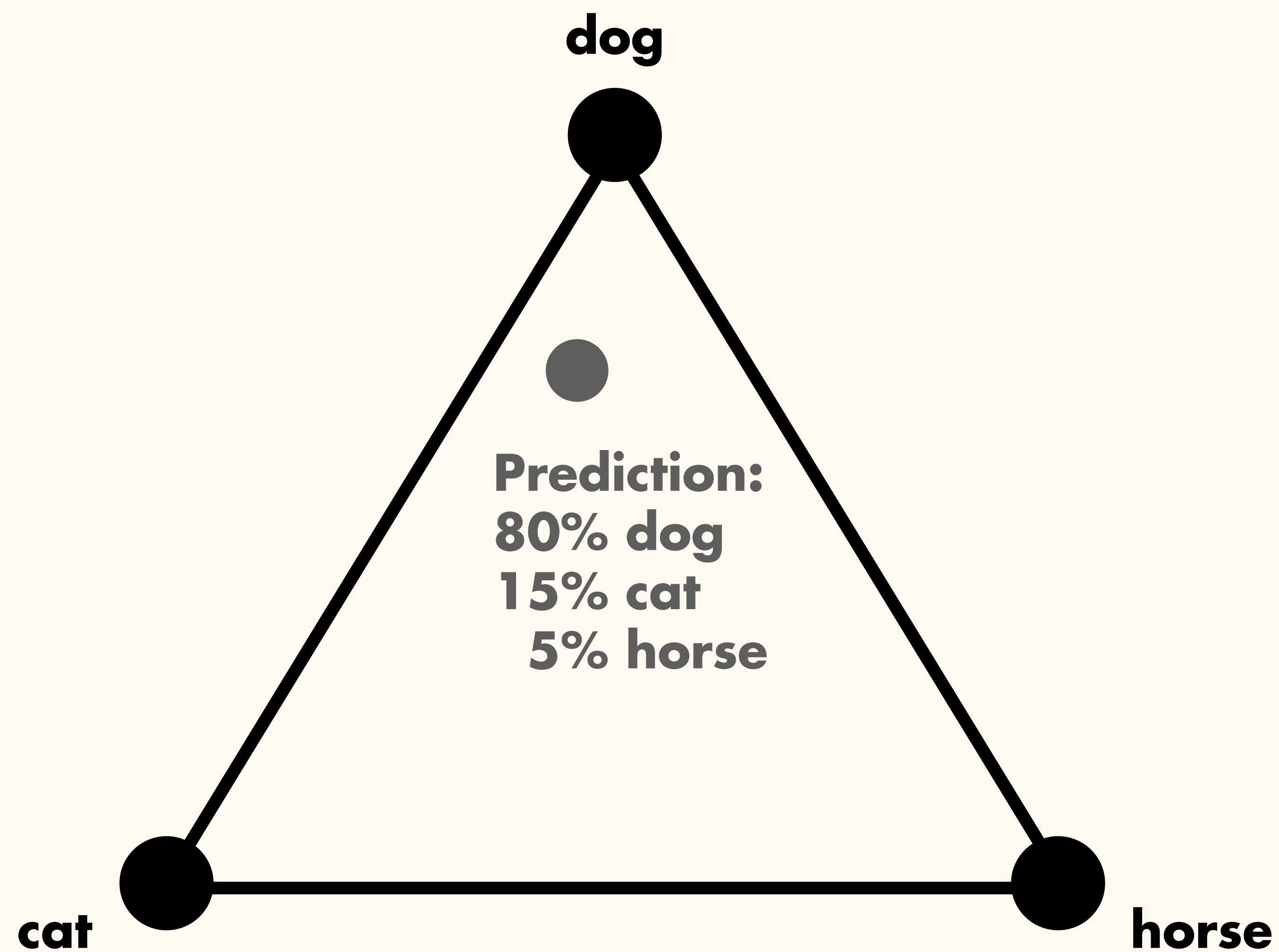




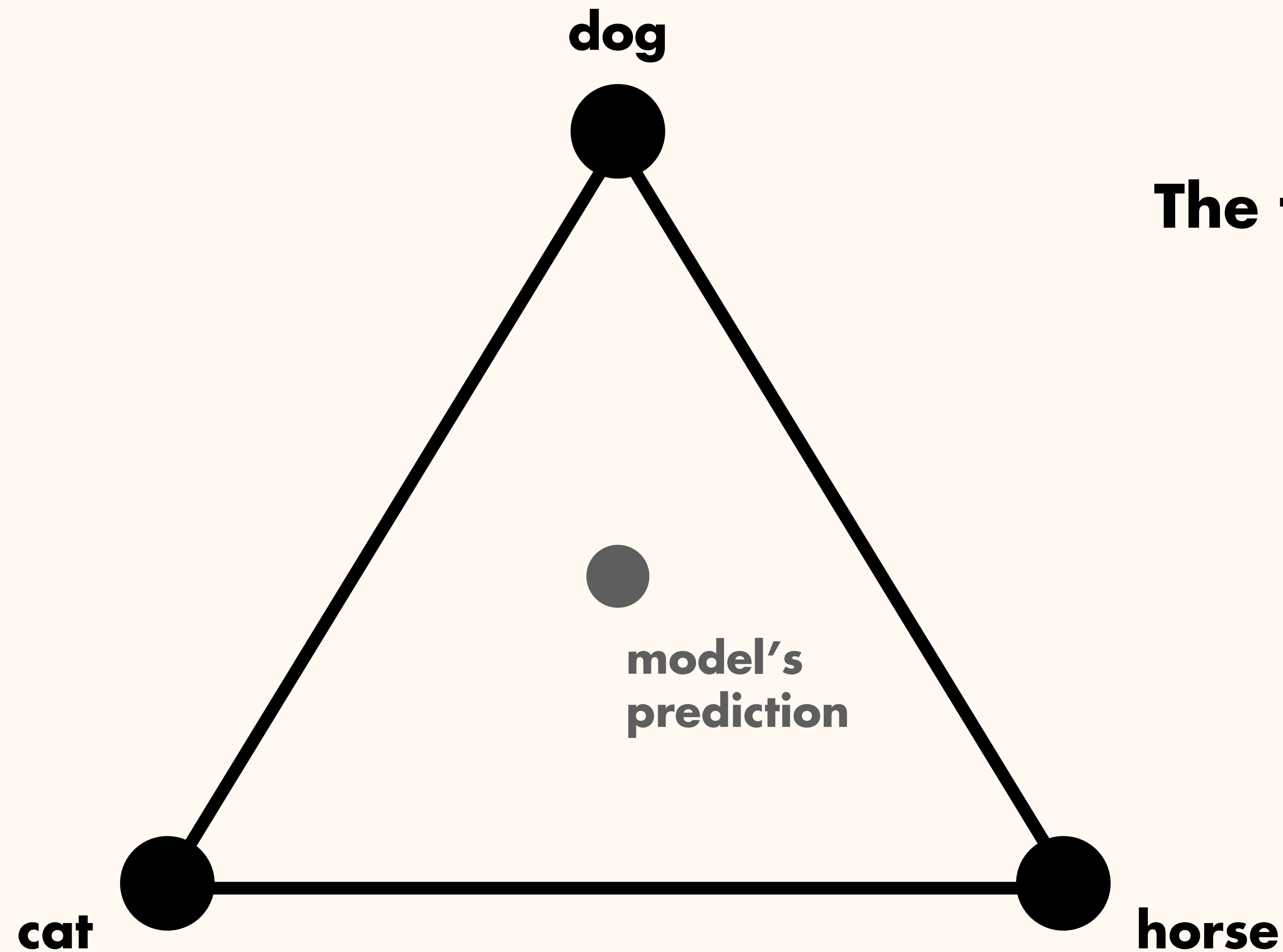






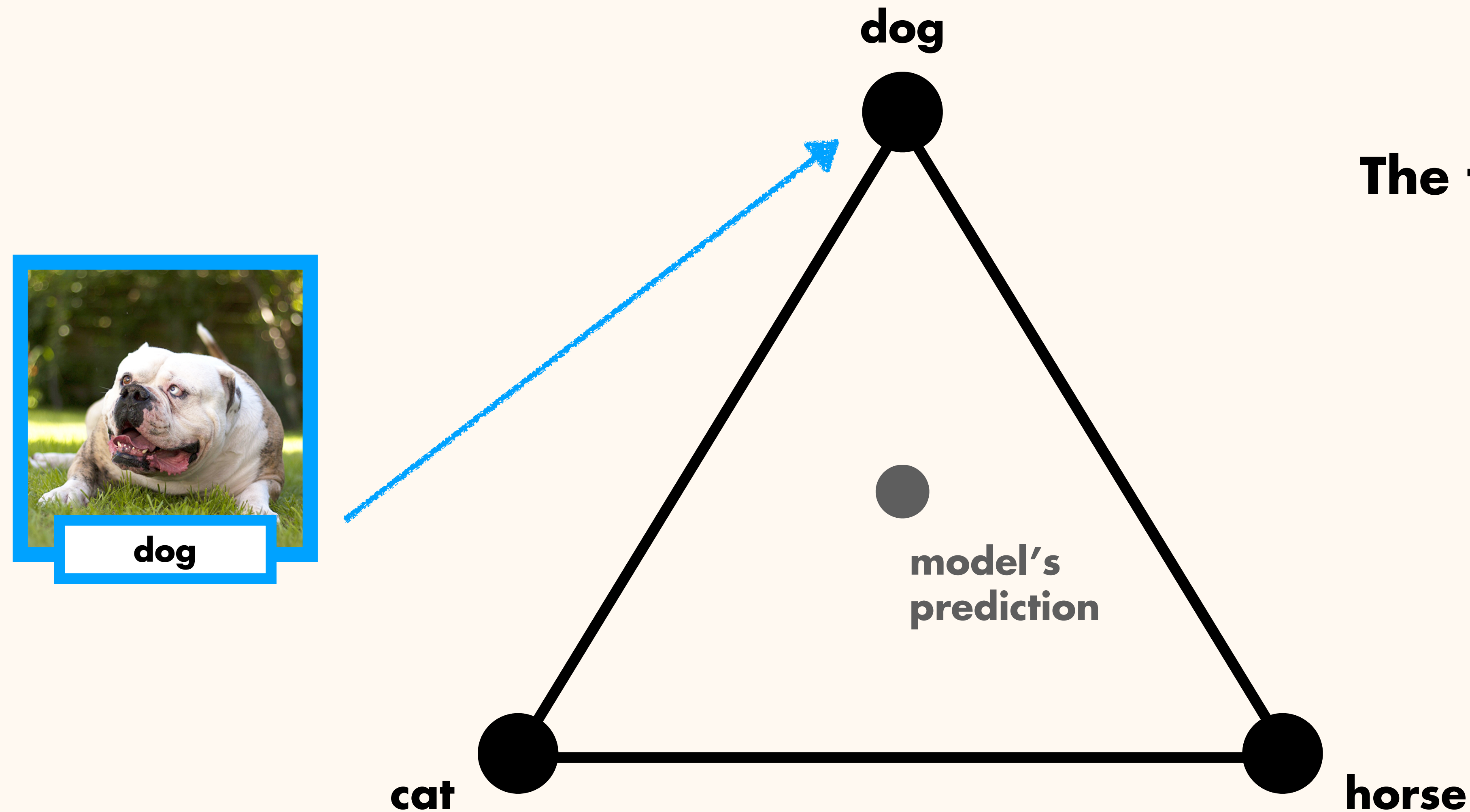


SUPERVISED LEARNING



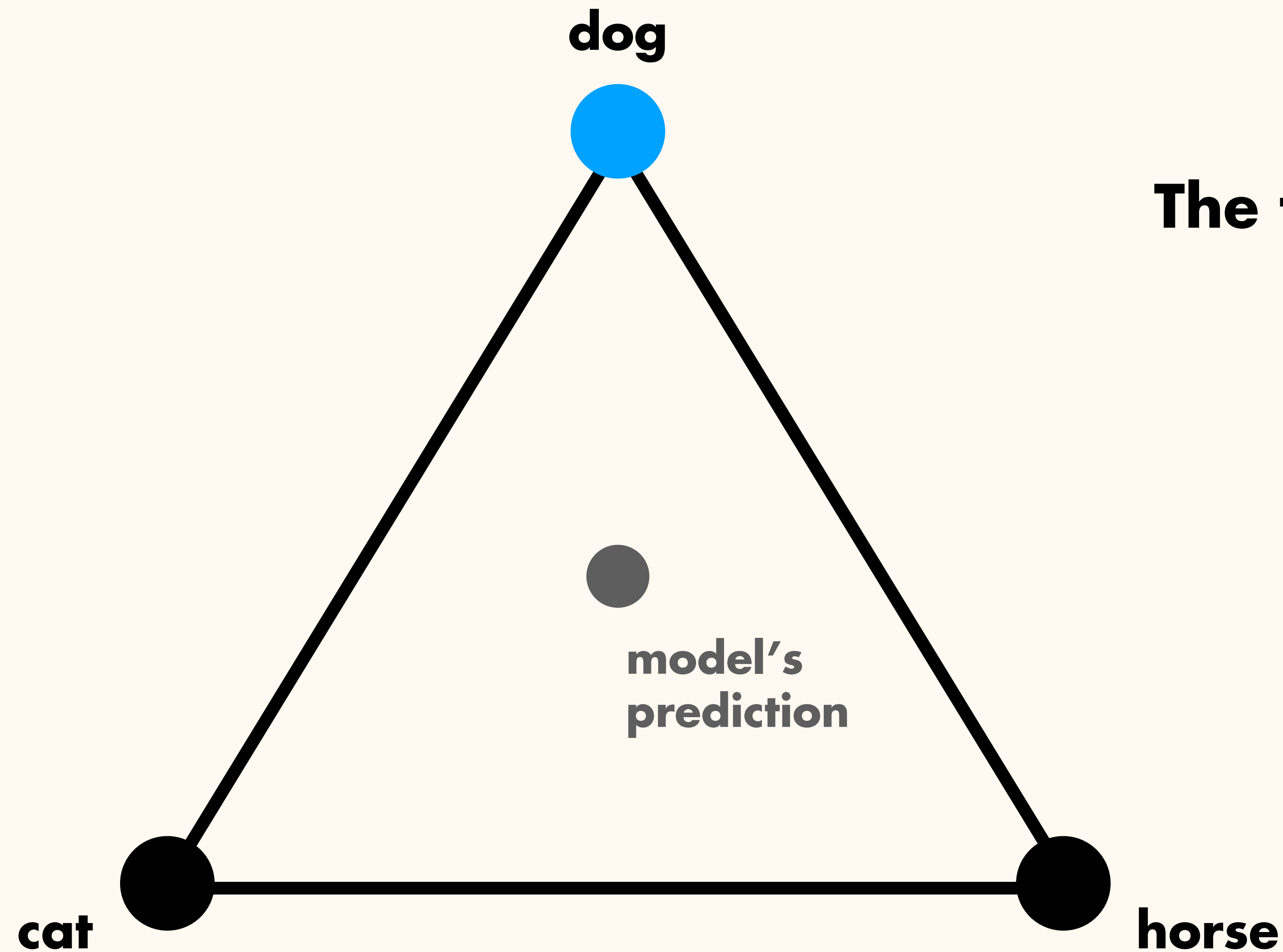
The true label pulls predictions to its direction.

SUPERVISED LEARNING

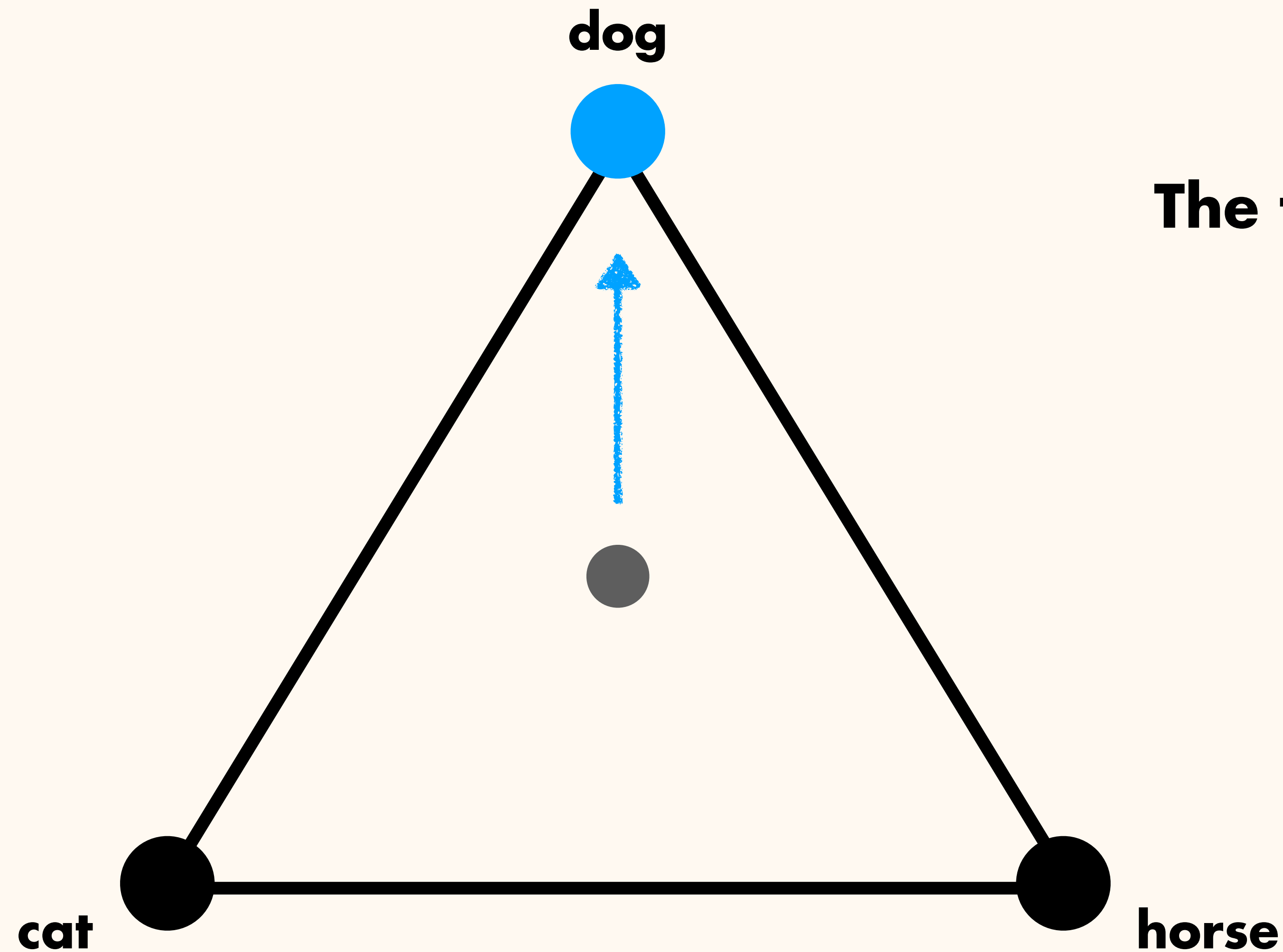


**The true label pulls
predictions to
its direction.**

SUPERVISED LEARNING

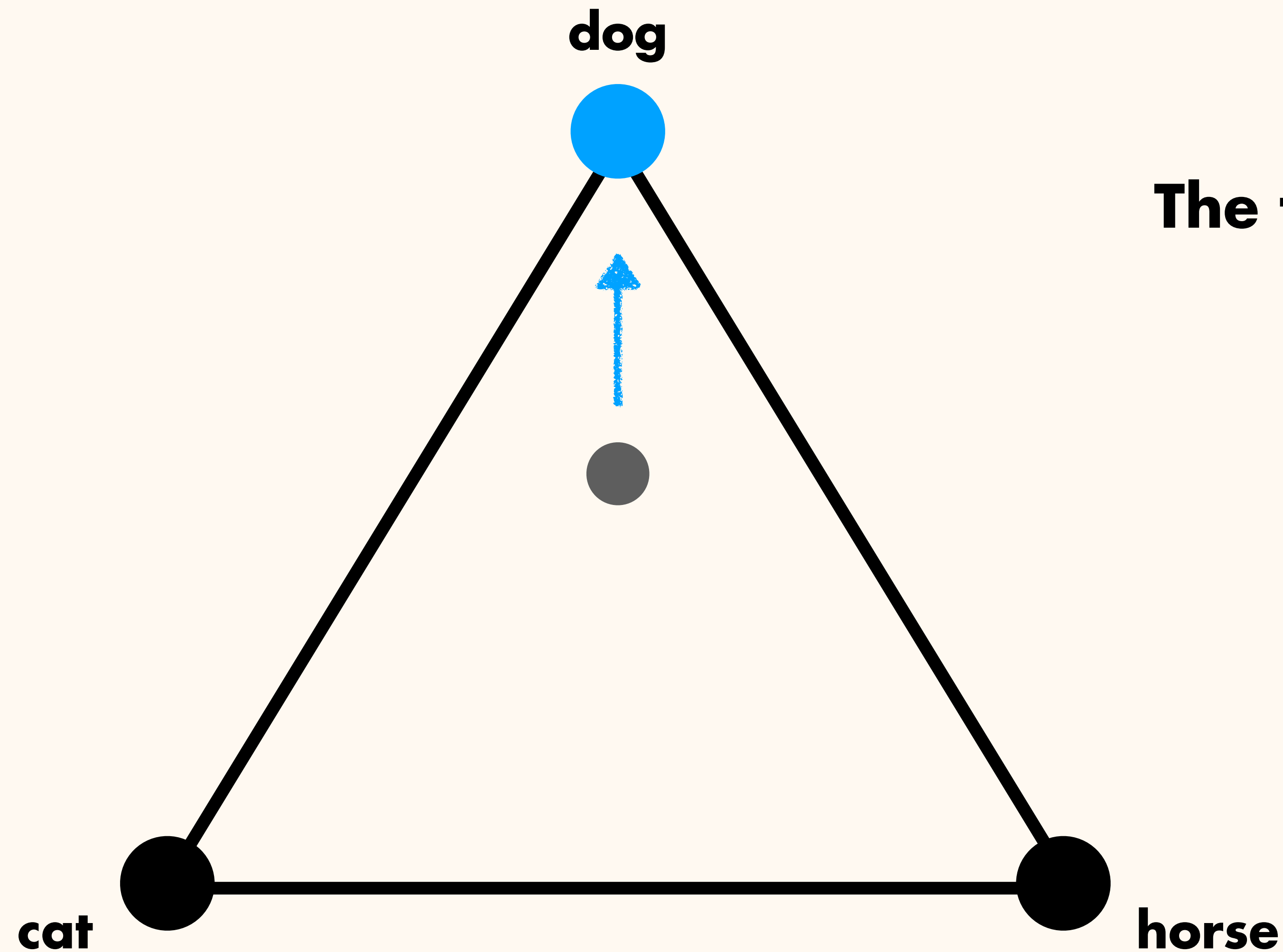


SUPERVISED LEARNING



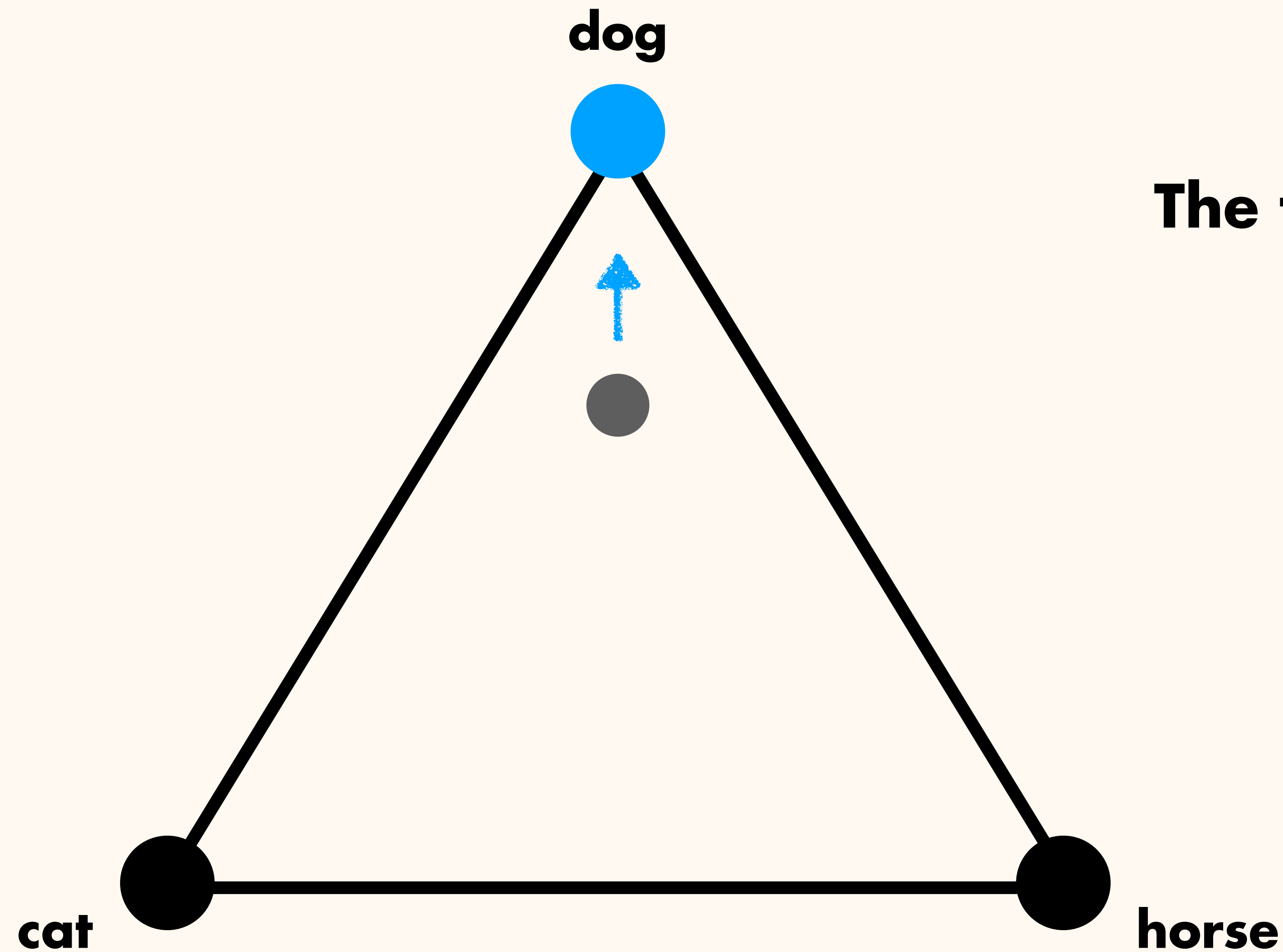
The true label pulls predictions to its direction.

SUPERVISED LEARNING



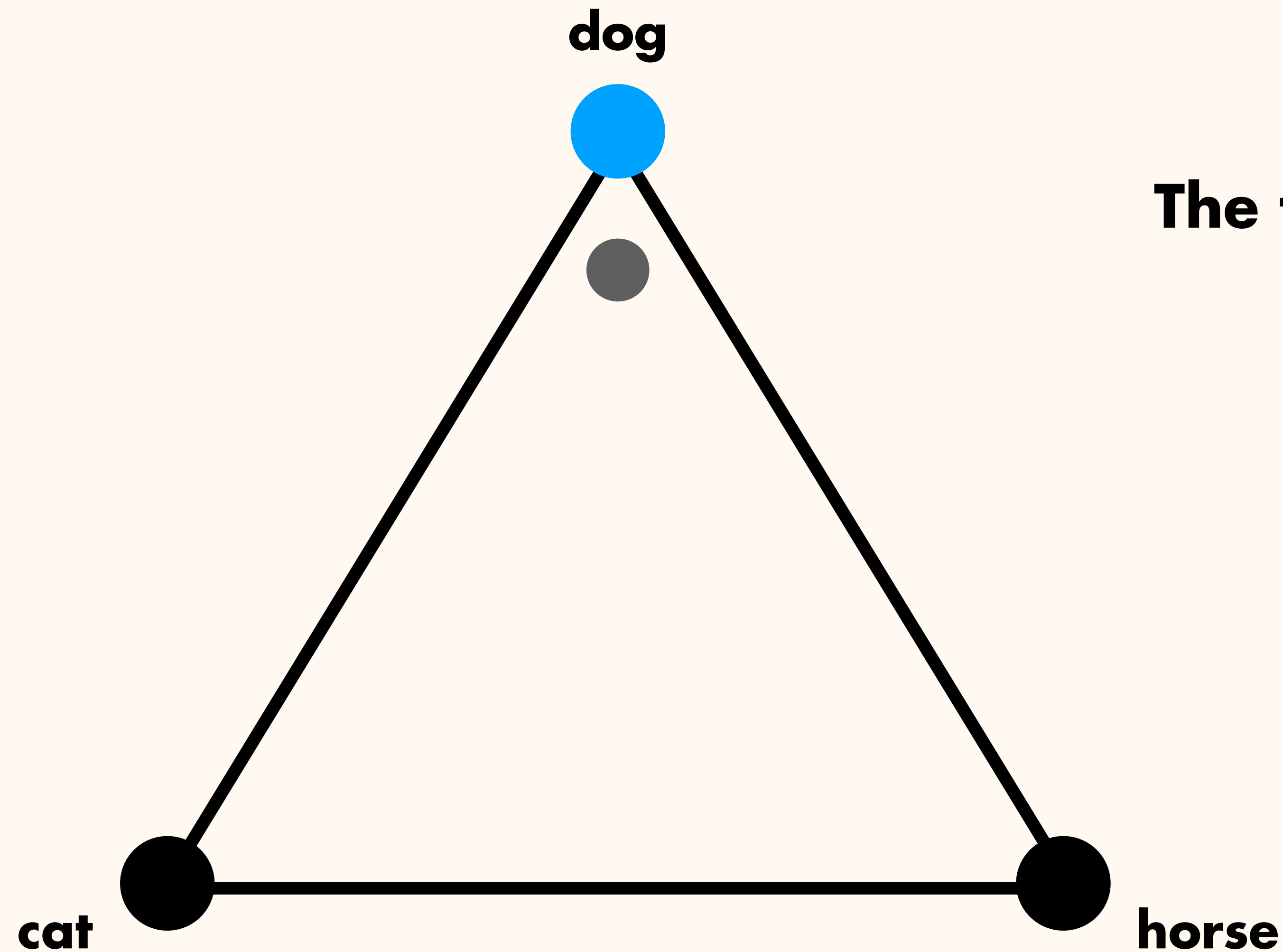
**The true label pulls
predictions to
its direction.**

SUPERVISED LEARNING



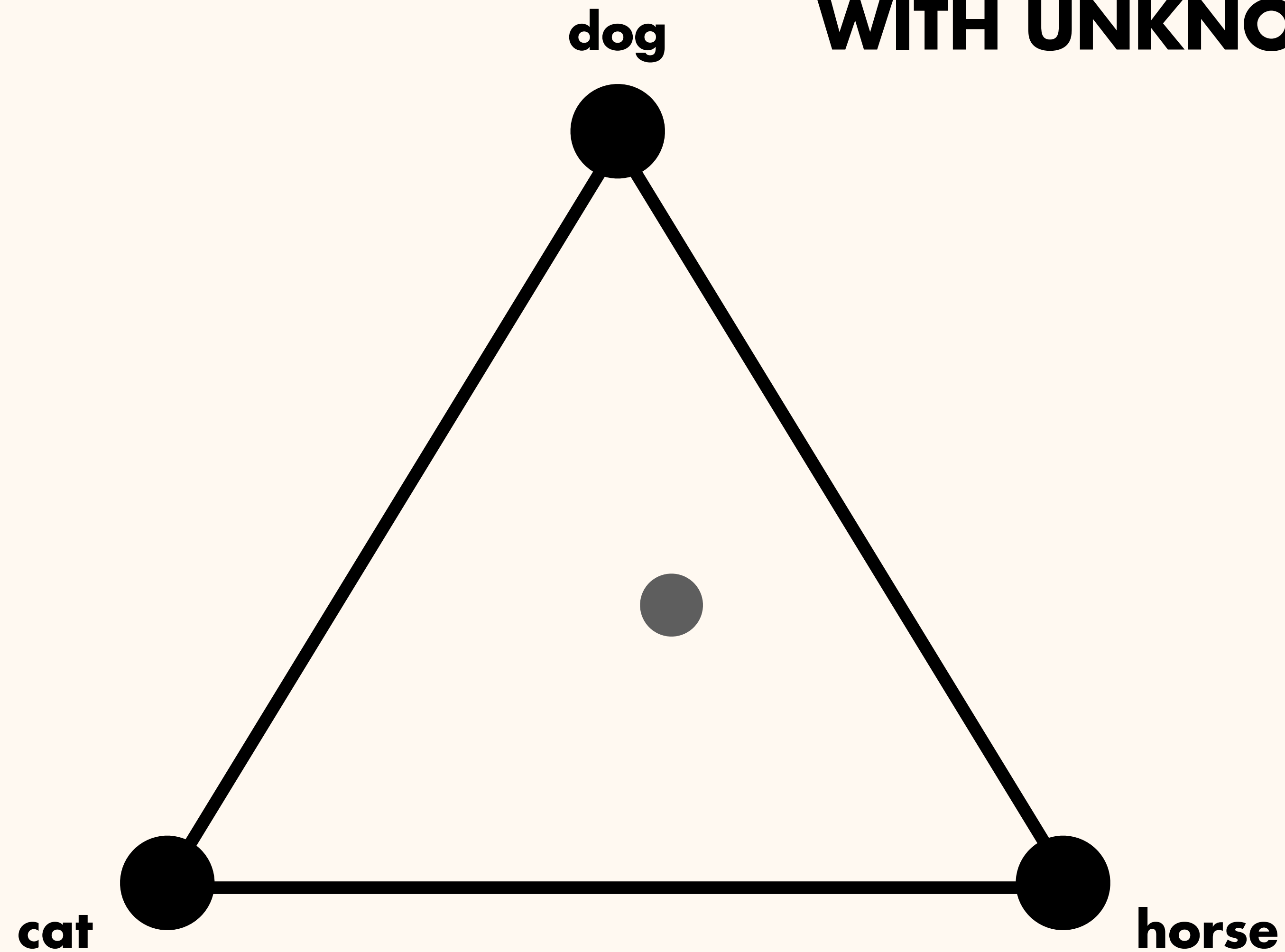
**The true label pulls
predictions to
its direction.**

SUPERVISED LEARNING



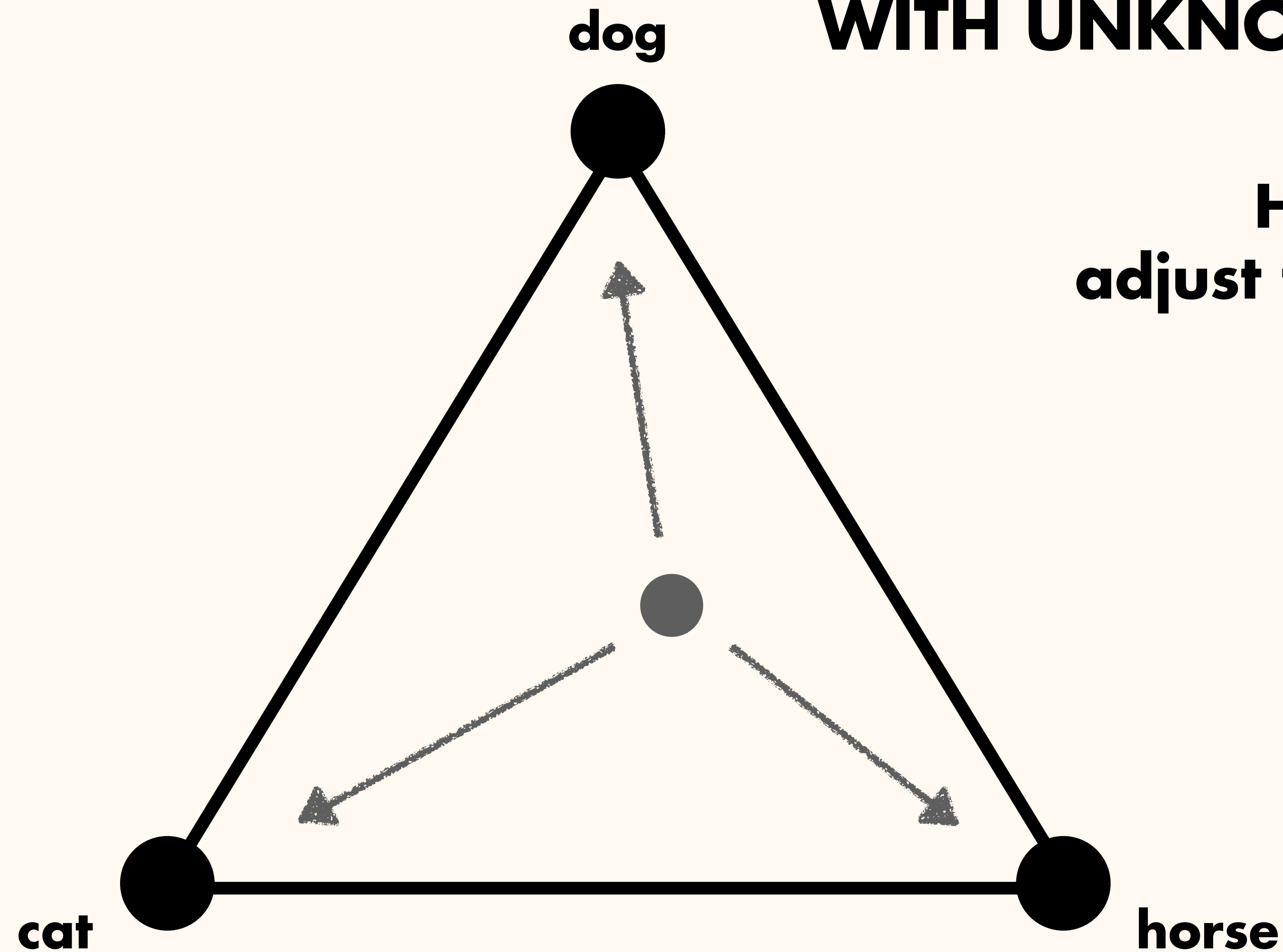
**The true label pulls
predictions to
its direction.**

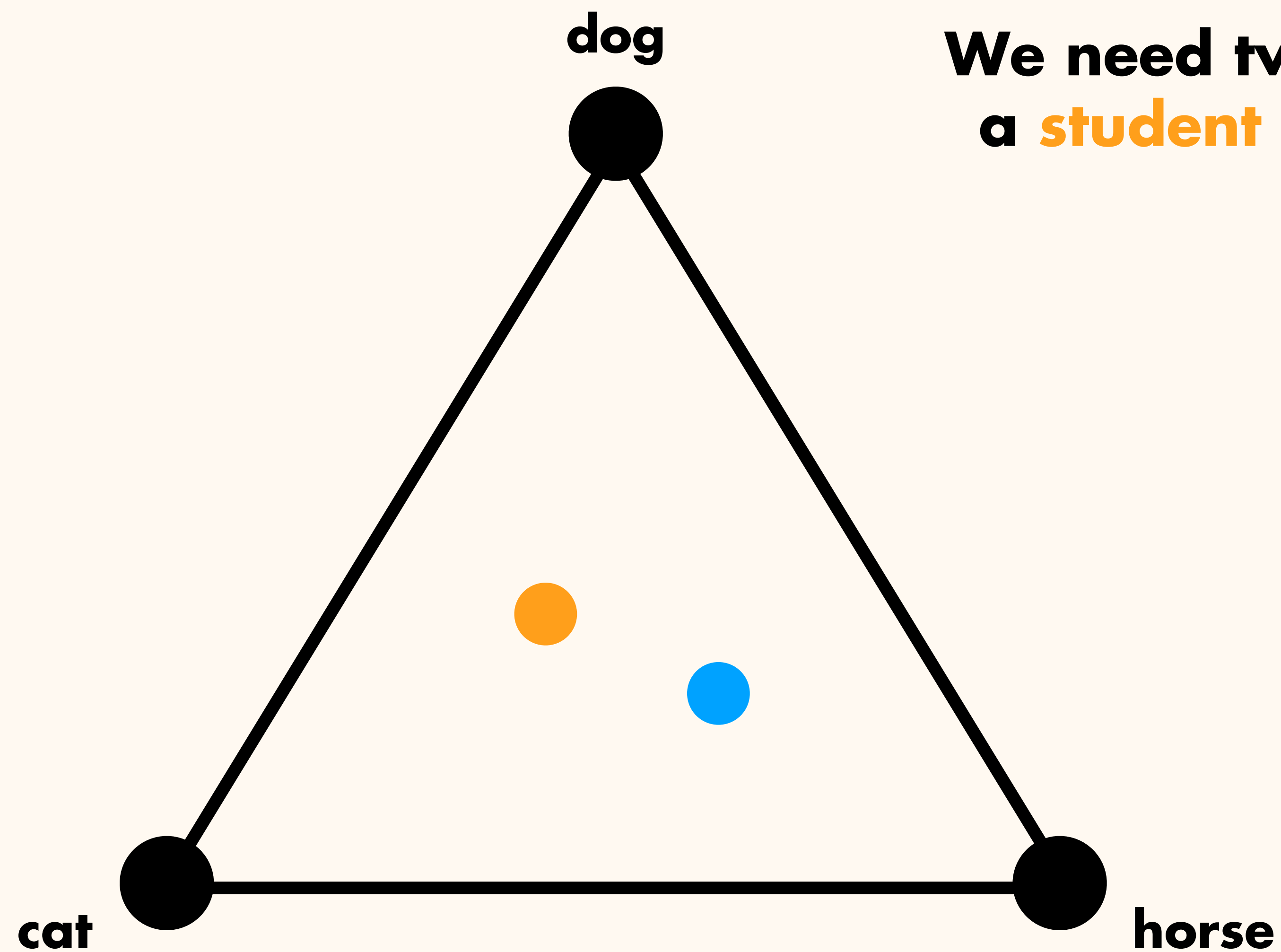
WHAT ABOUT EXAMPLES WITH UNKNOWN LABELS?



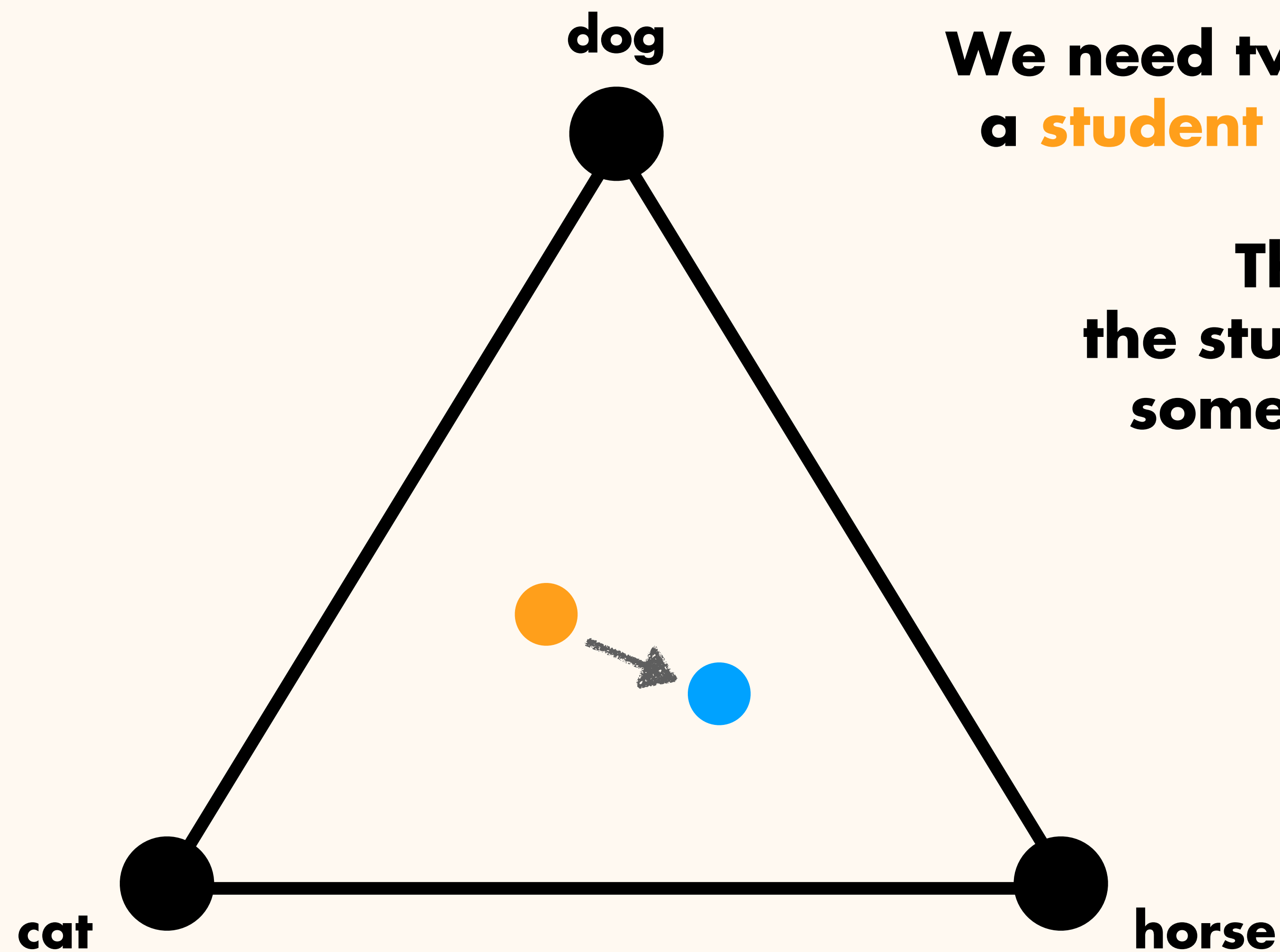
WHAT ABOUT EXAMPLES WITH UNKNOWN LABELS?

How should we adjust the prediction?





We need two predictions:
a **student** and a **teacher**.

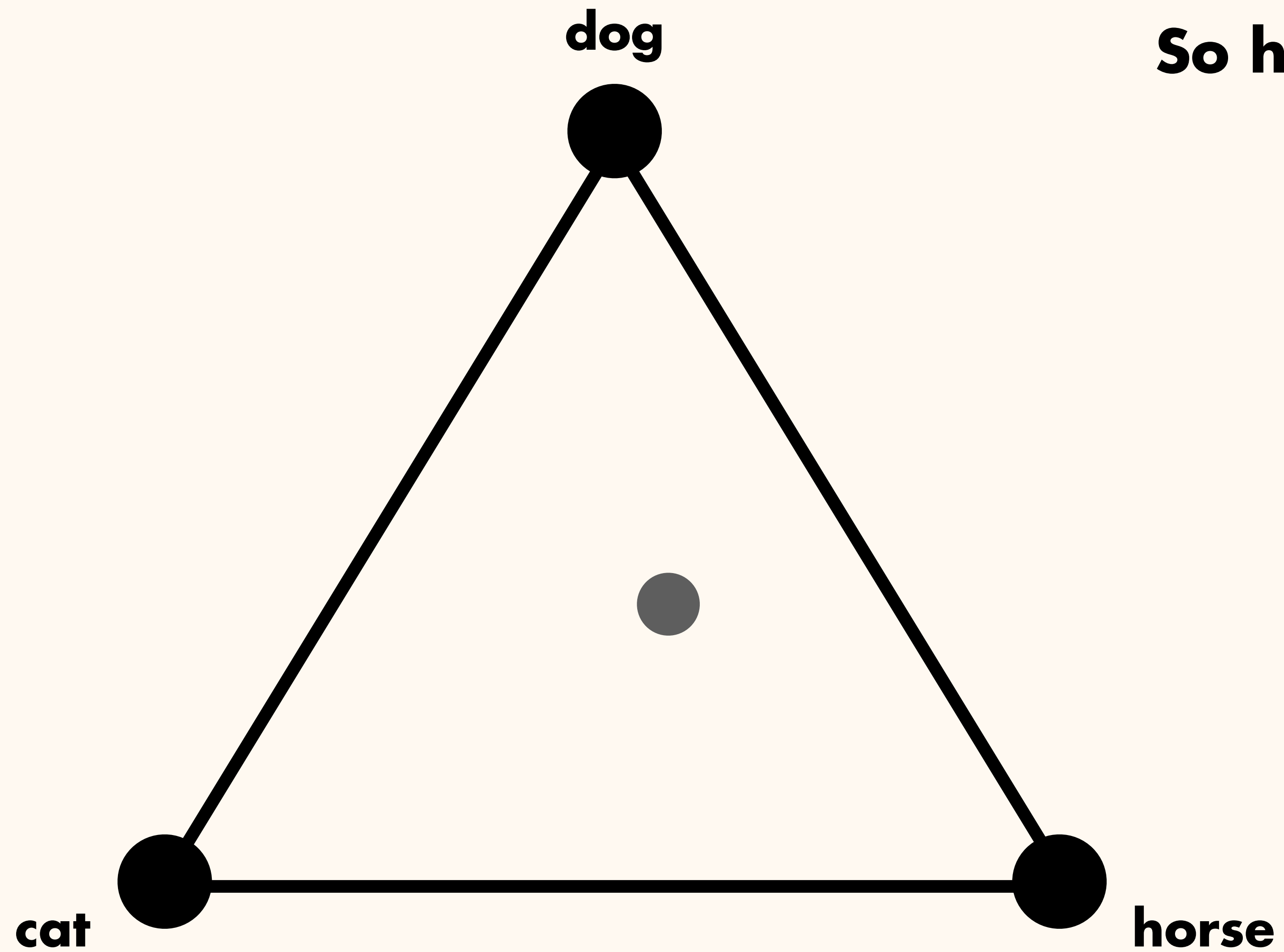


We need two predictions:
a **student** and a **teacher**.

Then, hopefully,
the student can learn
something from the
teacher.



???

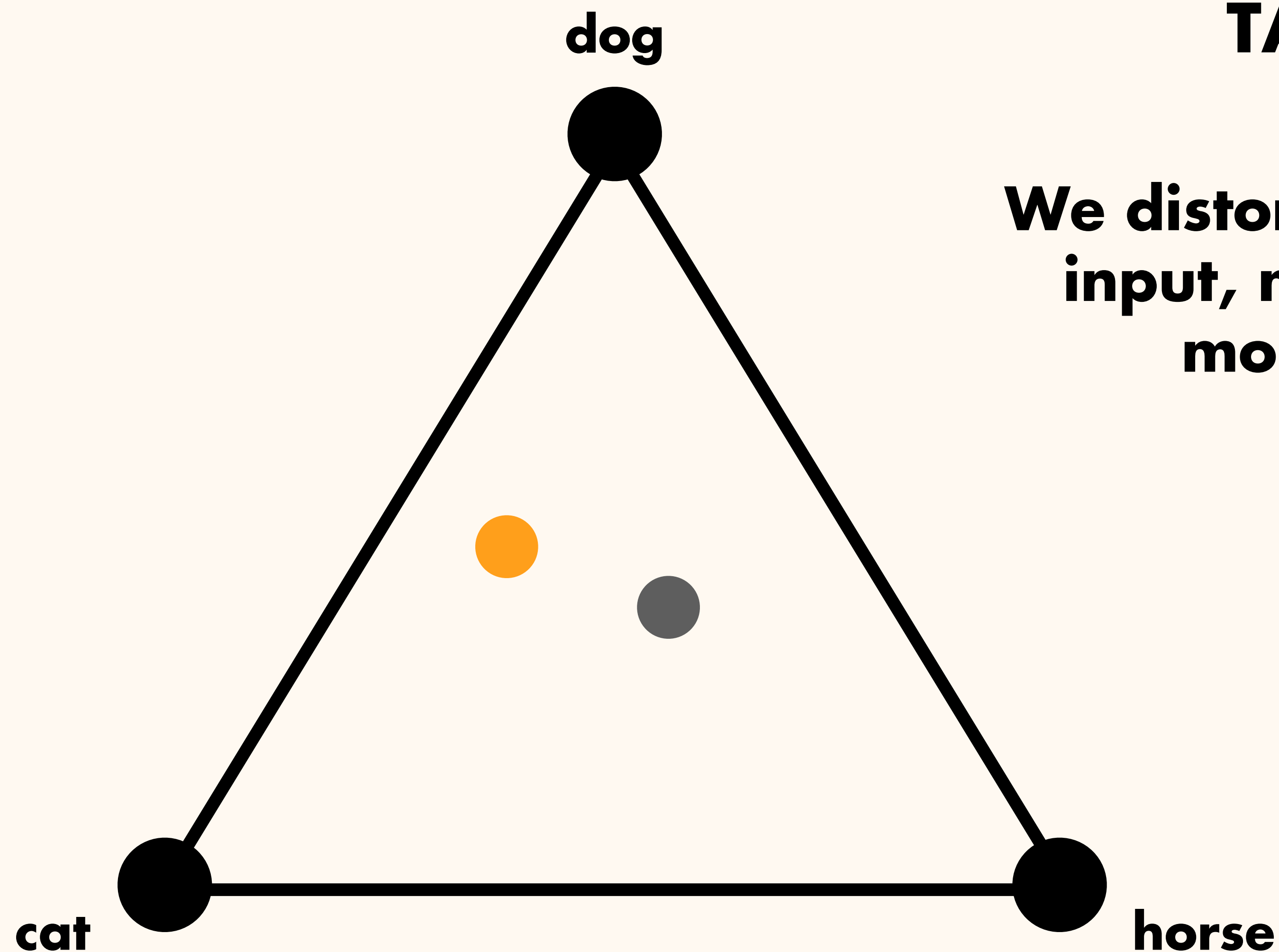


So how to do this?

Two ways.

1. MAKE THE TASK HARDER.

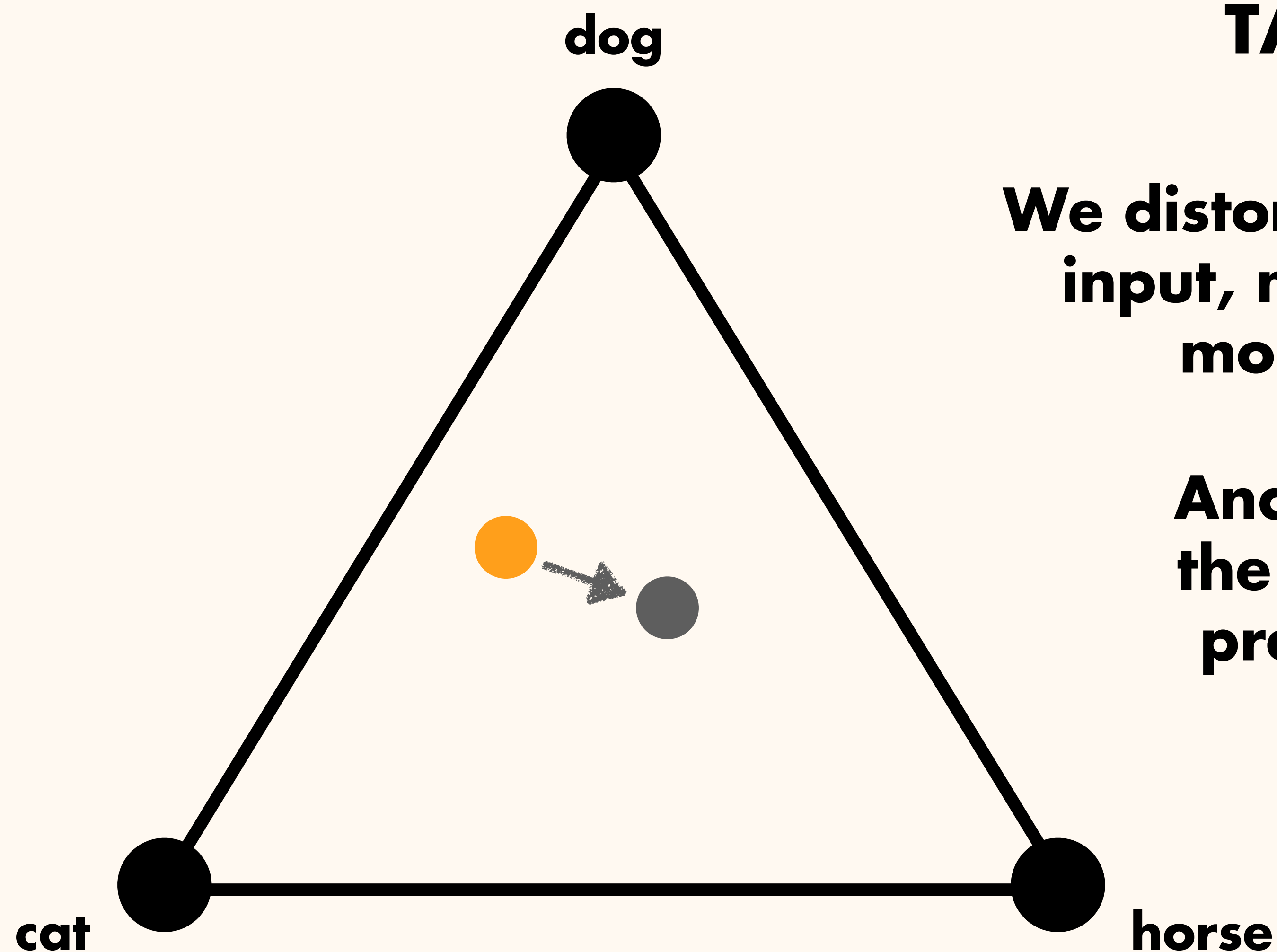
We distort the **student's** input, making its task more challenging.



1. MAKE THE TASK HARDER.

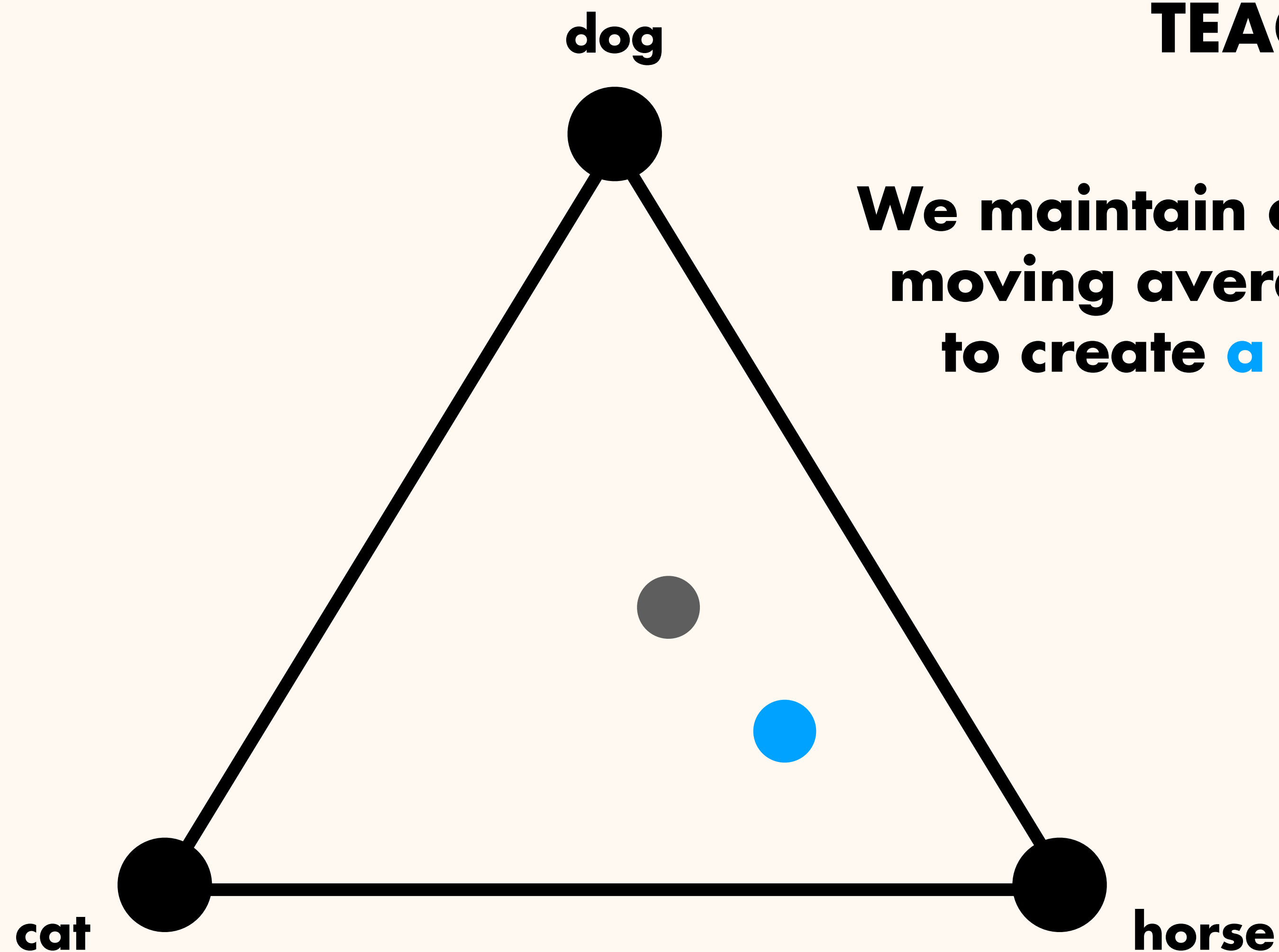
We distort the **student's** input, making its task more challenging.

And then we train the harder task to predict the easier tasks' output.



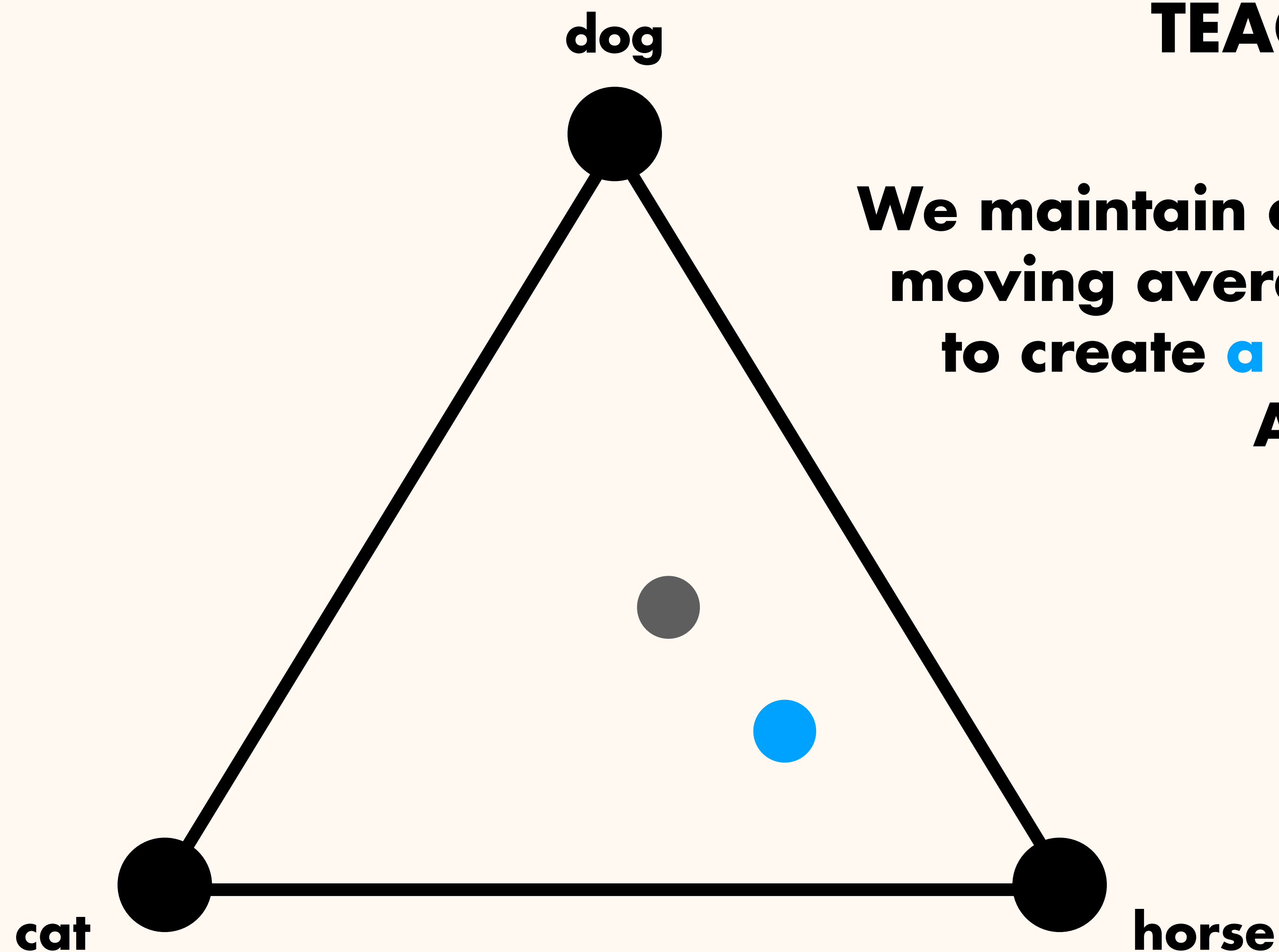
2. MAKE THE TEACHER BETTER.

We maintain an exponential
moving average of weights
to create **a better teacher.**



2. MAKE THE TEACHER BETTER.

We maintain an exponential
moving average of weights
to create **a better teacher.**
A mean teacher.

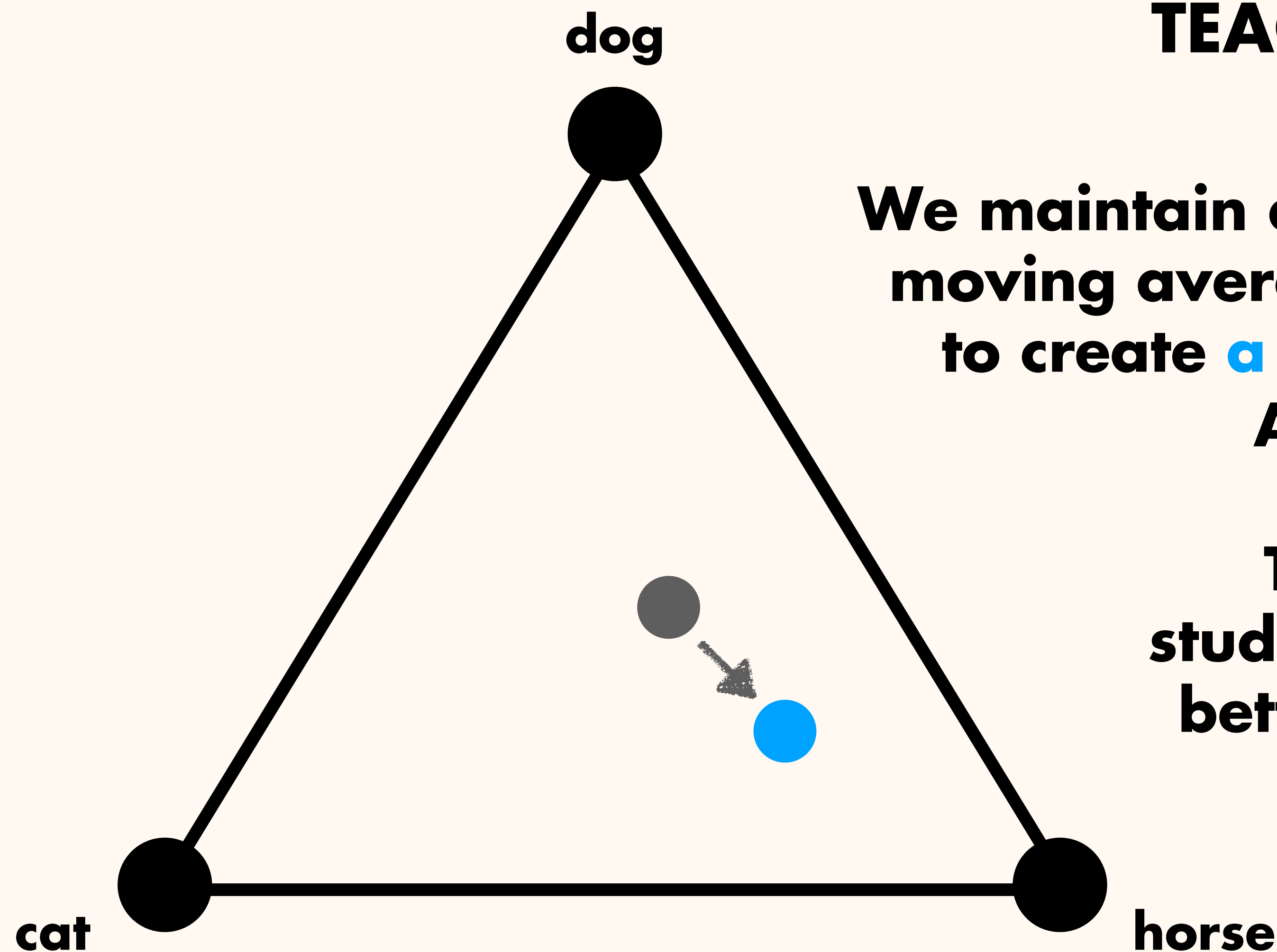


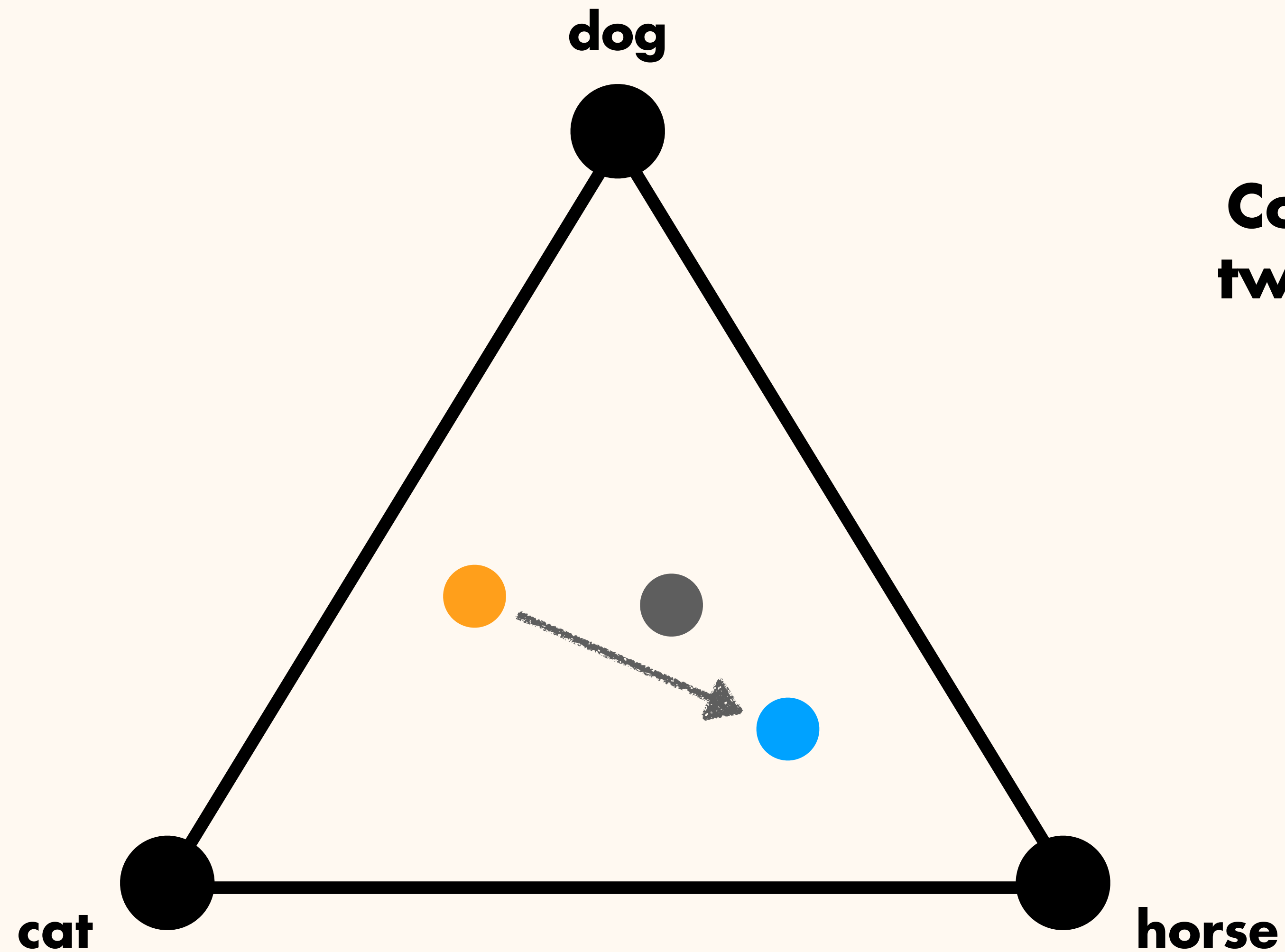
2. MAKE THE TEACHER BETTER.

We maintain an exponential
moving average of weights
to create **a better teacher.**

A mean teacher.

Then we let the
student learn these
better predictions.

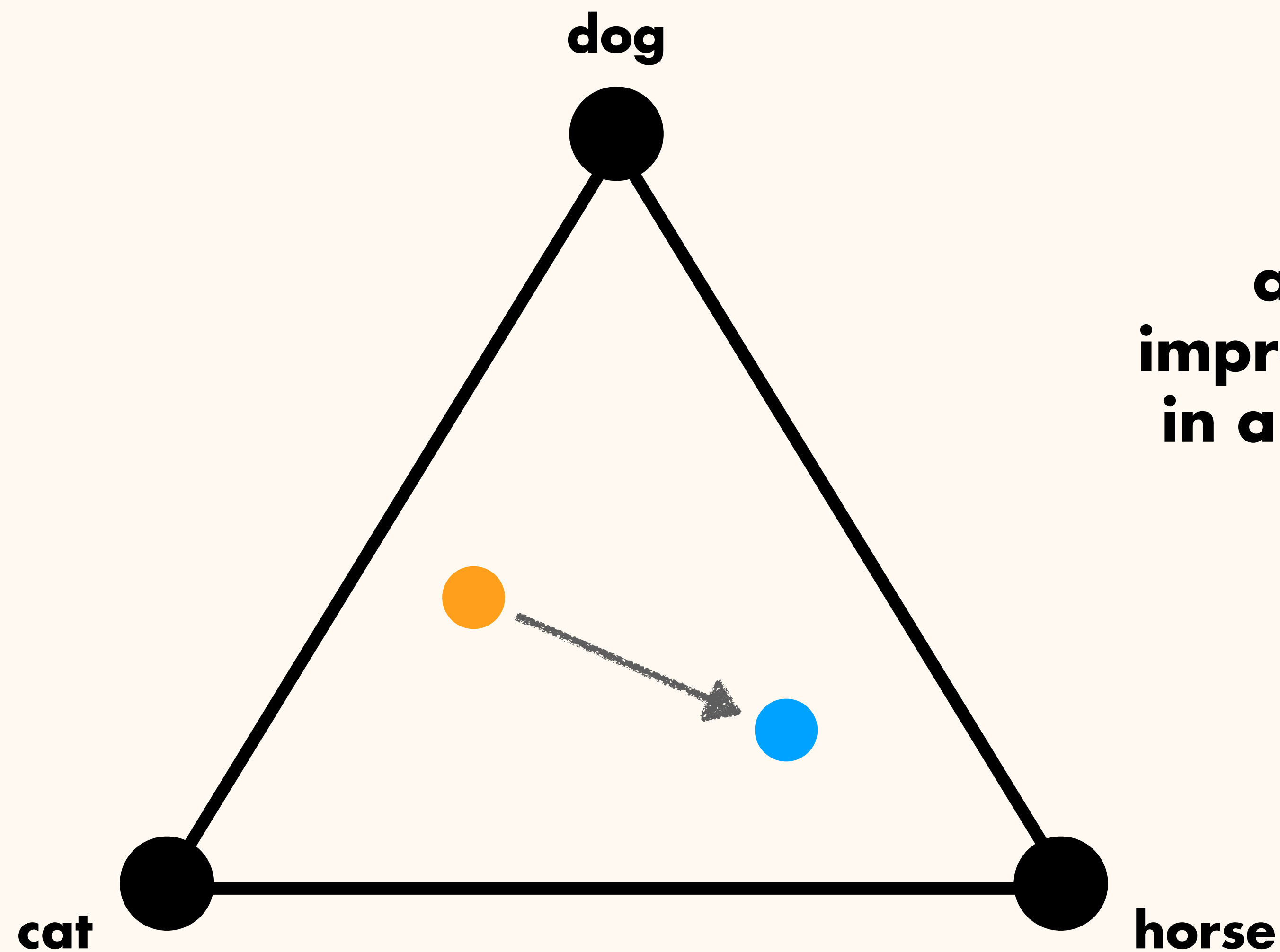




**Combining these
two ways works
even better.**



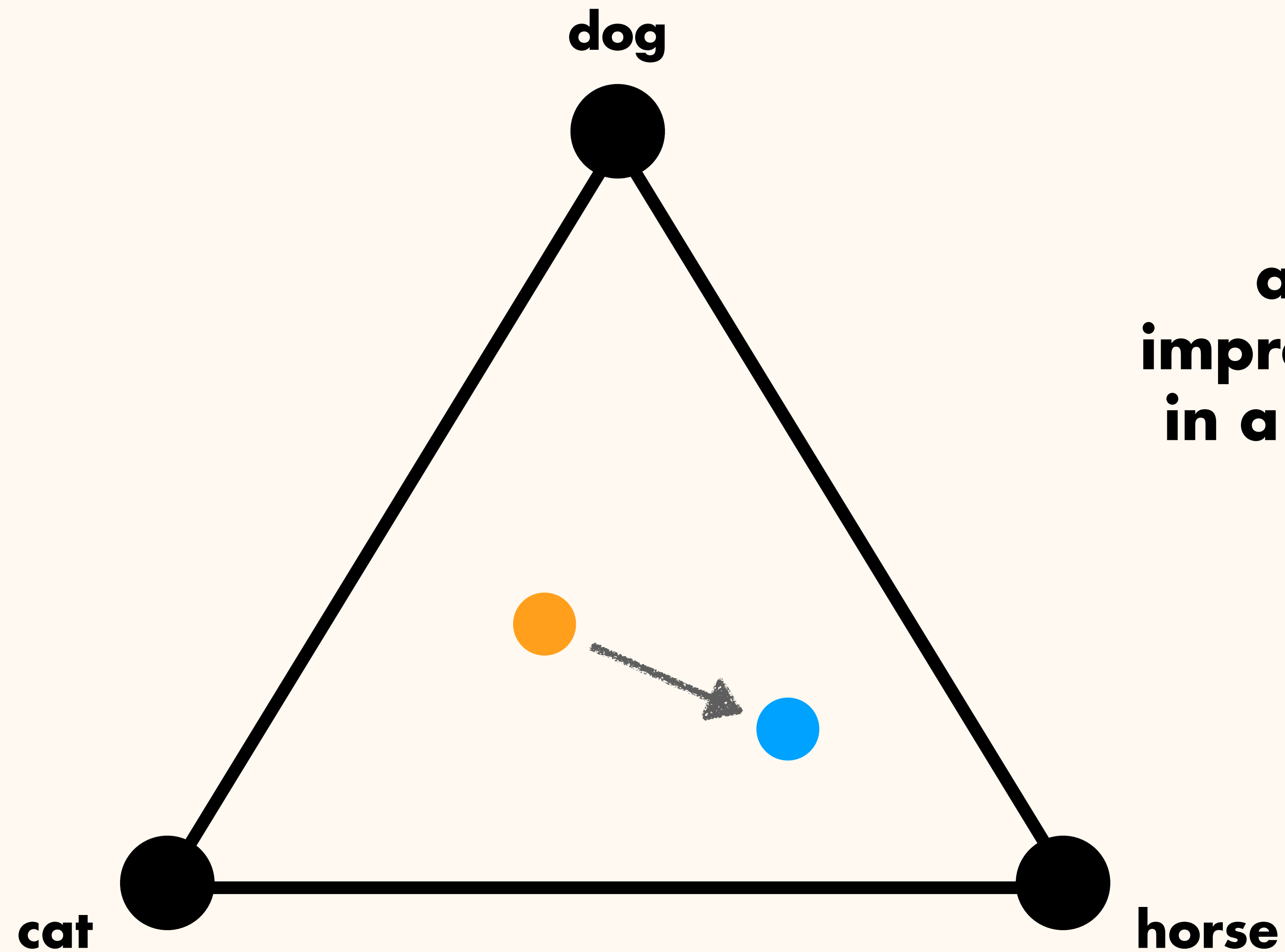
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



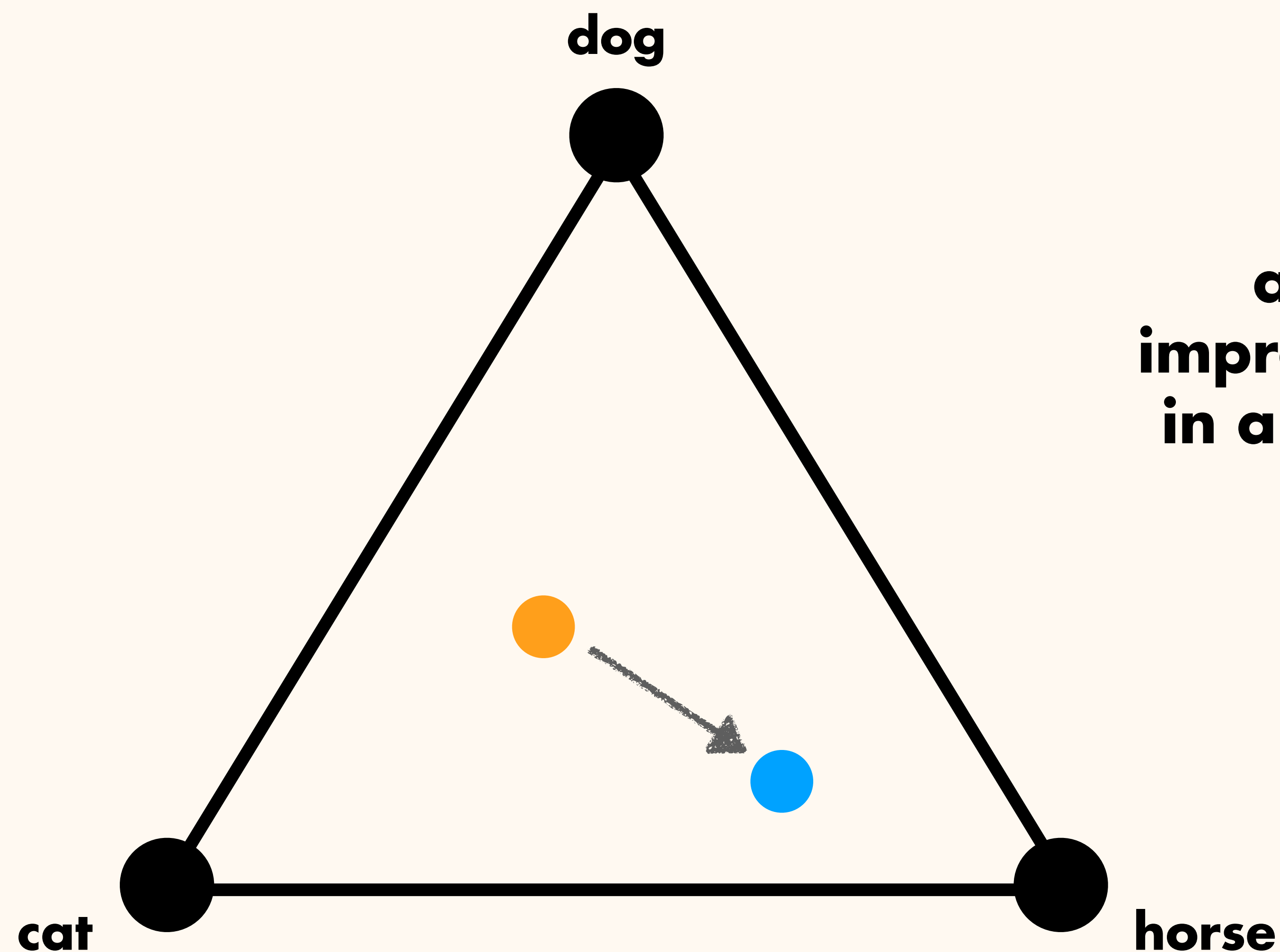
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



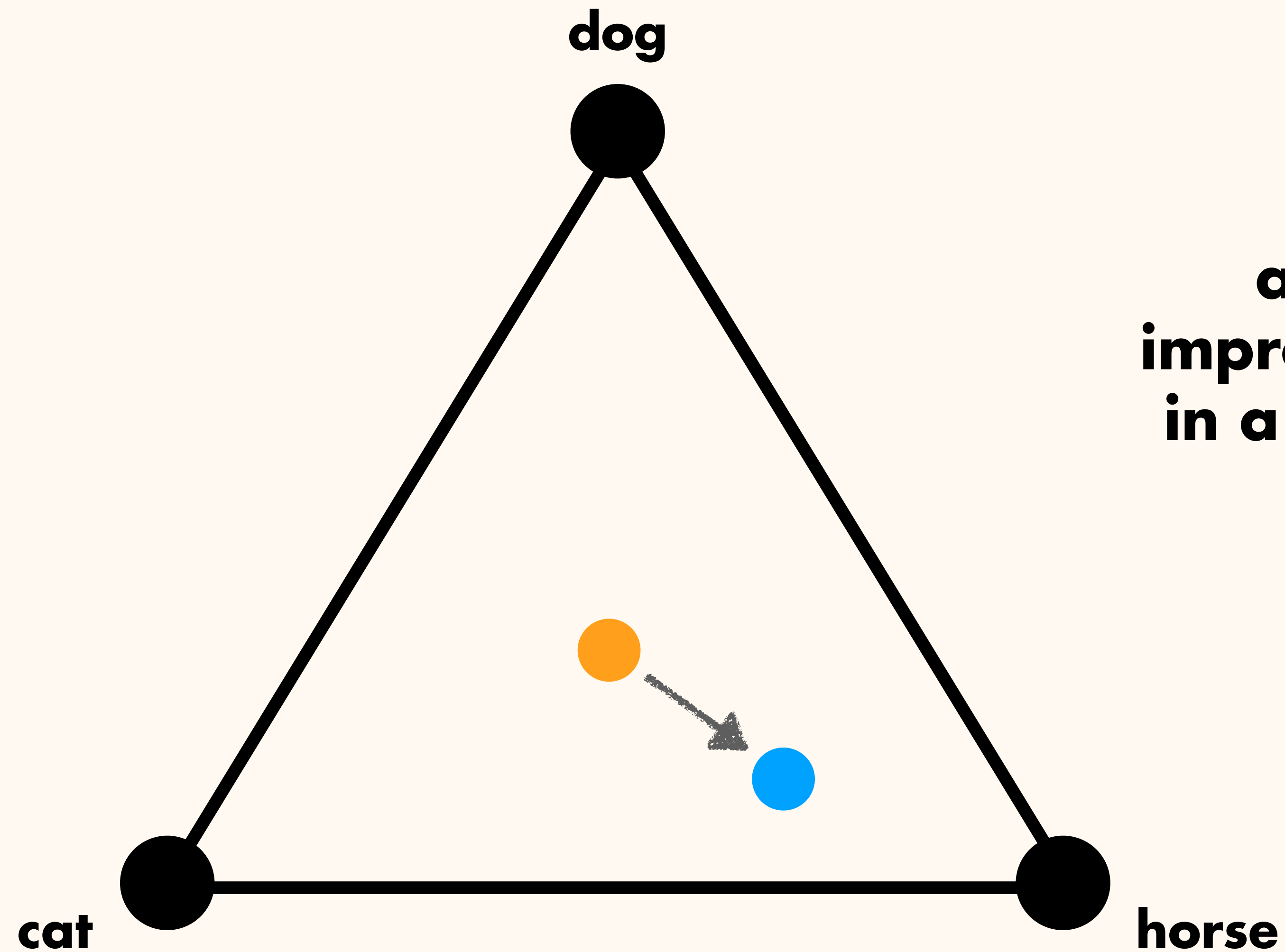
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



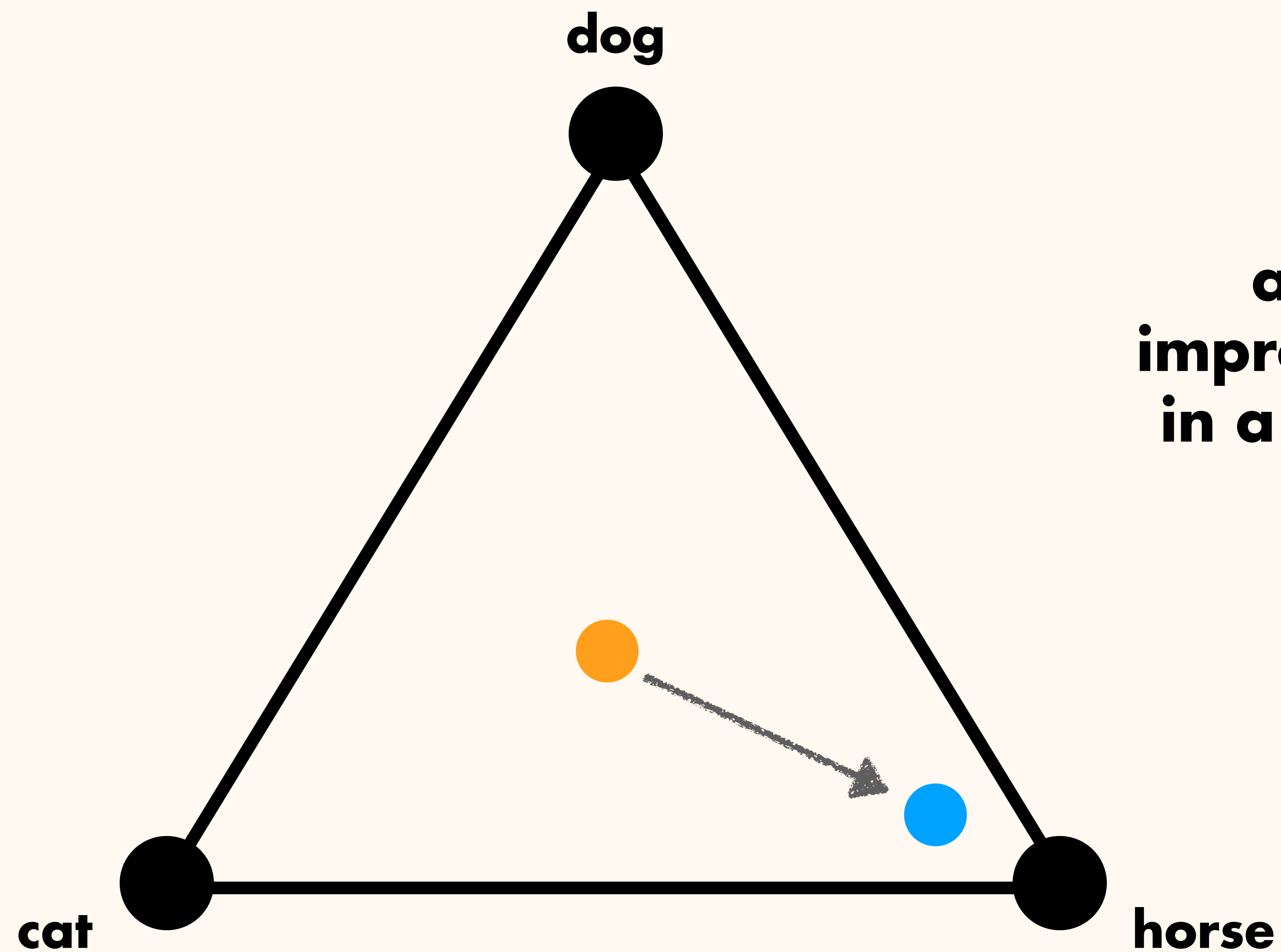
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



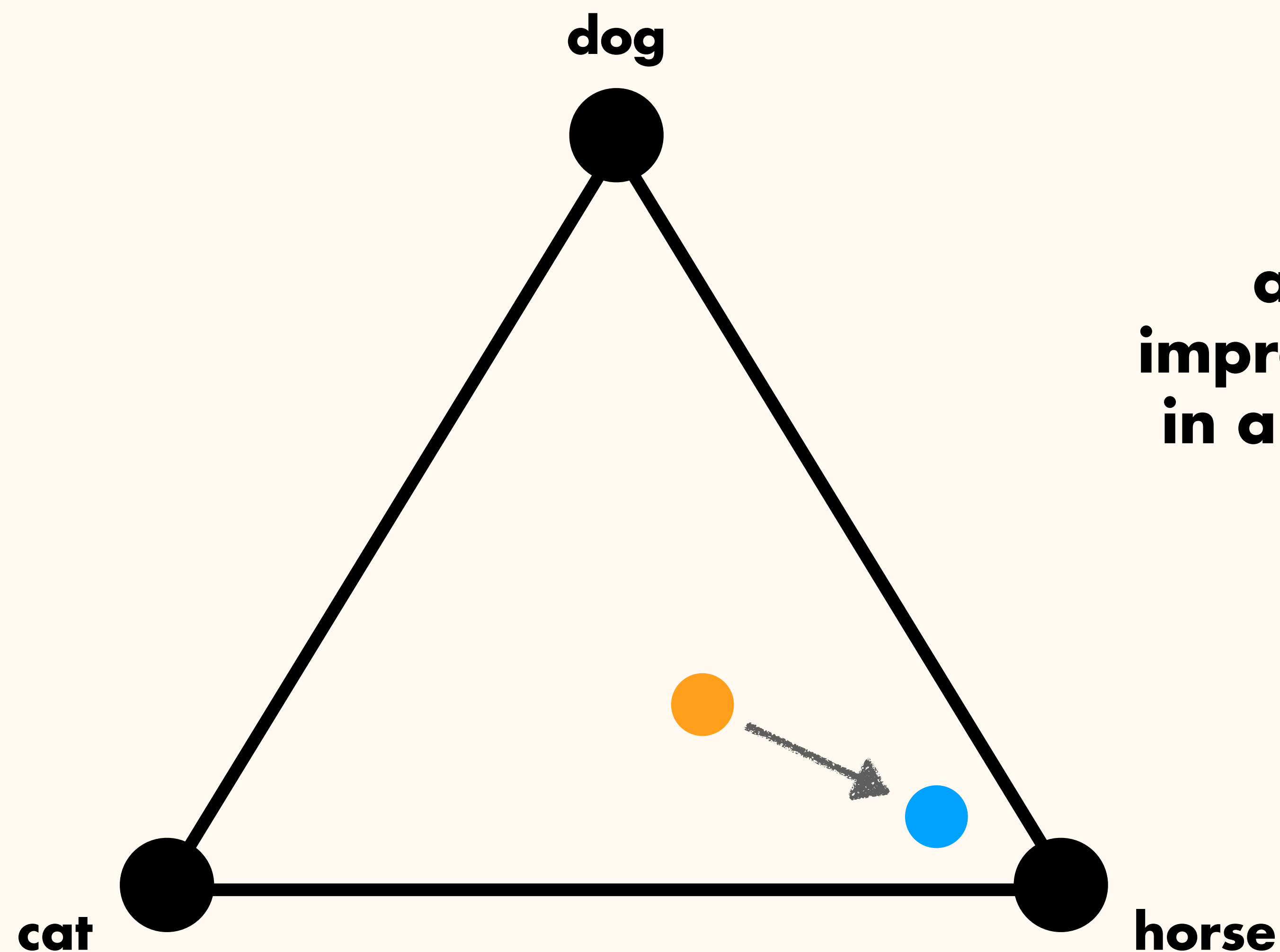
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



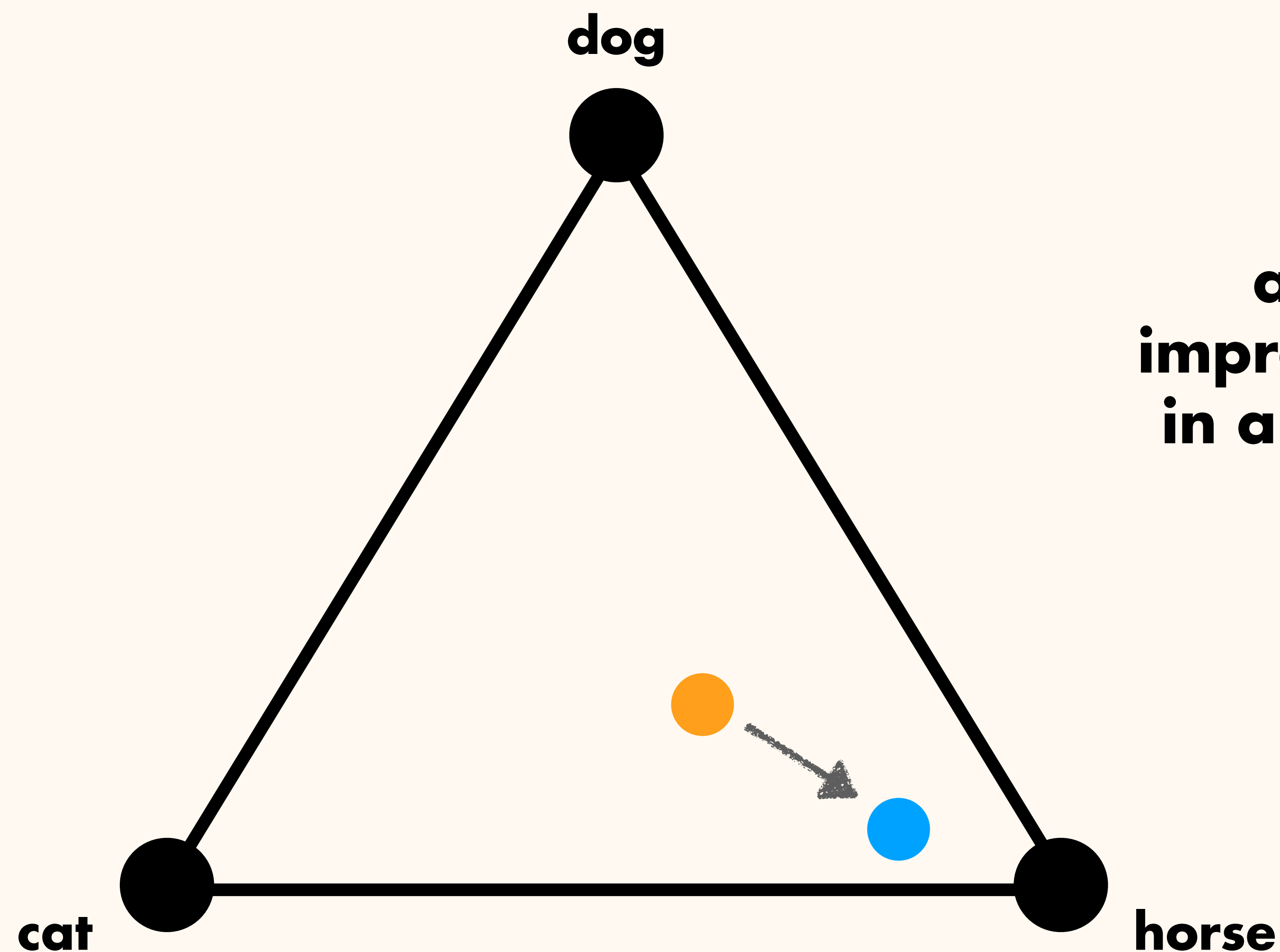
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



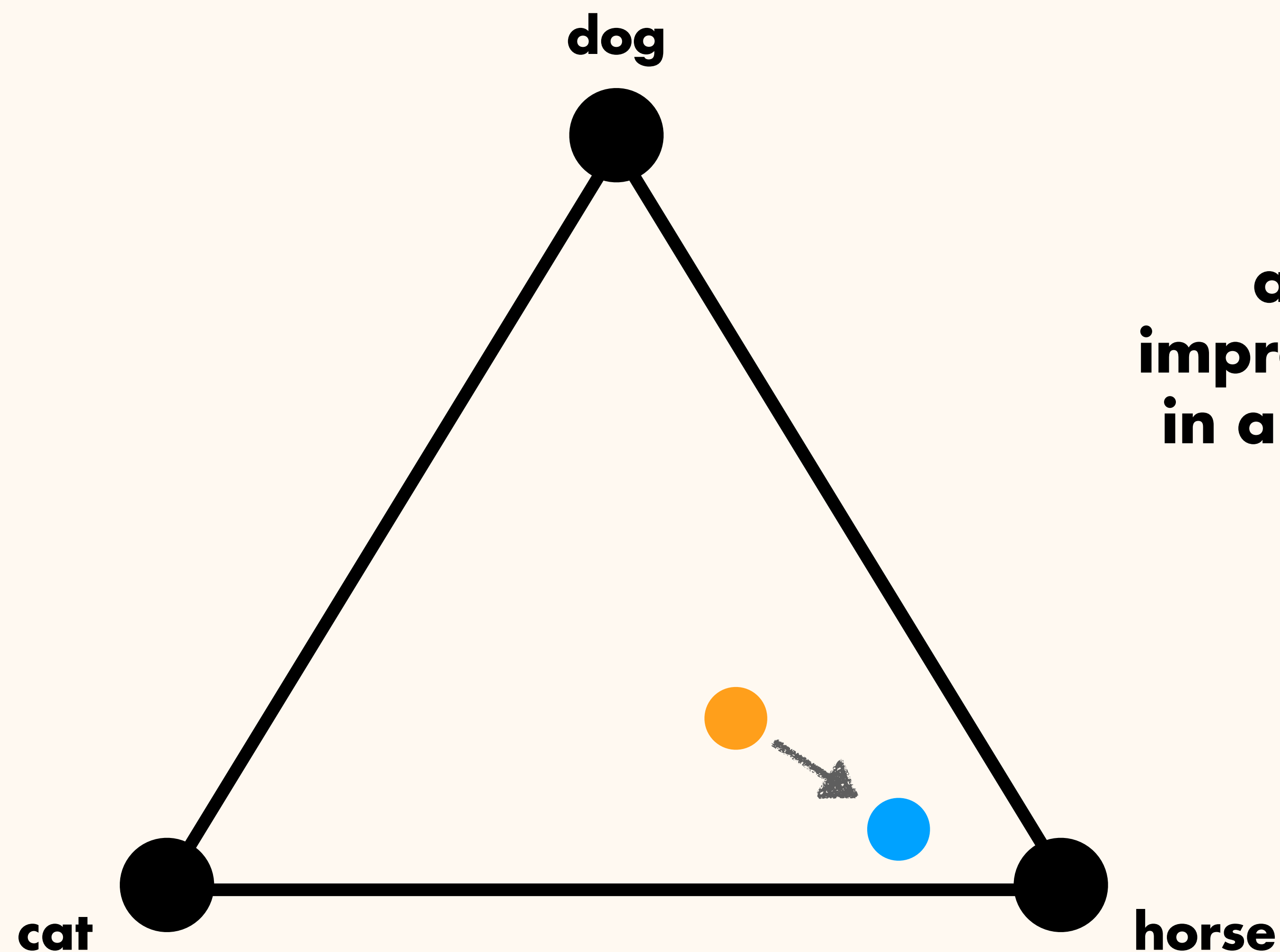
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



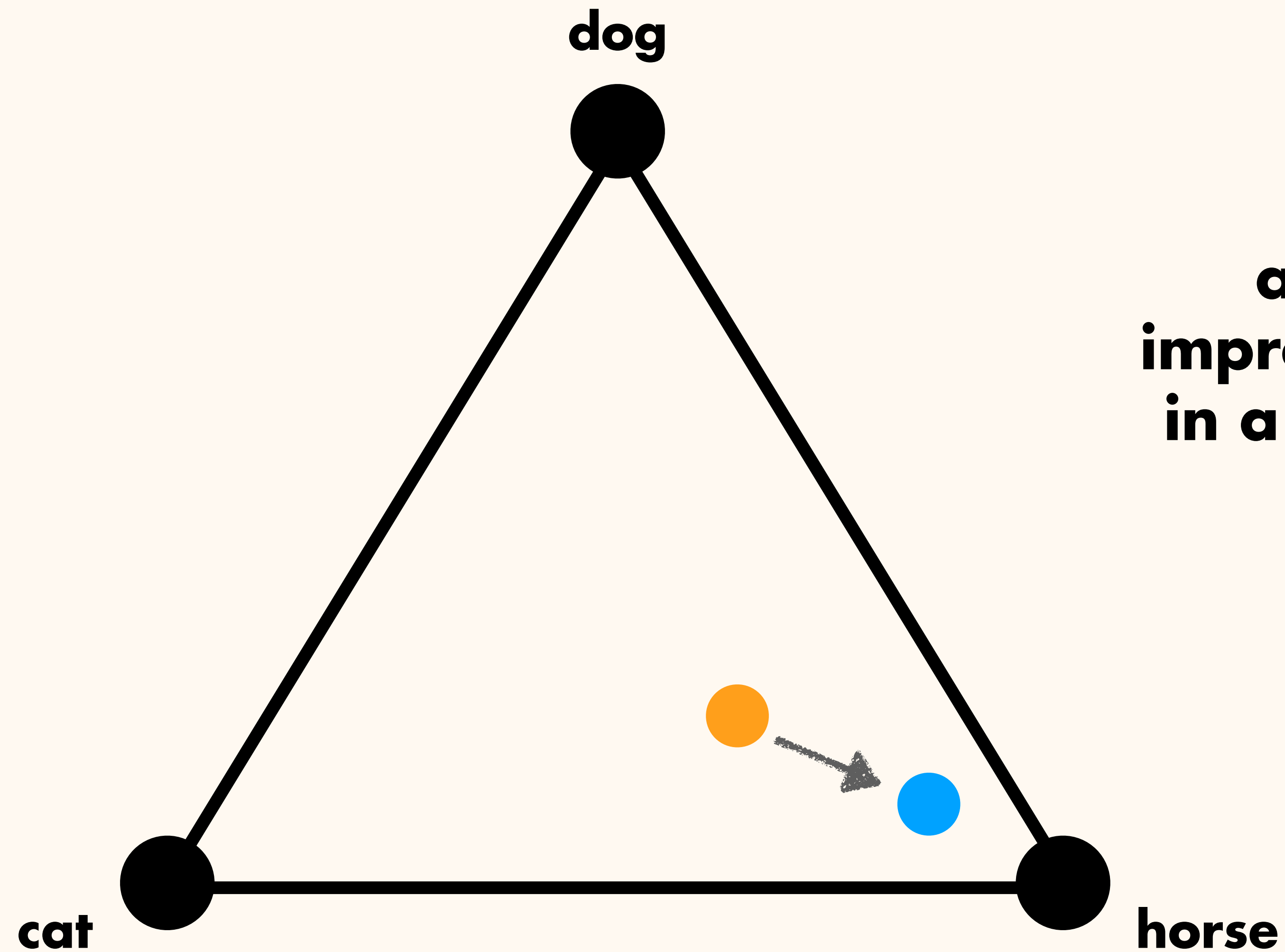
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



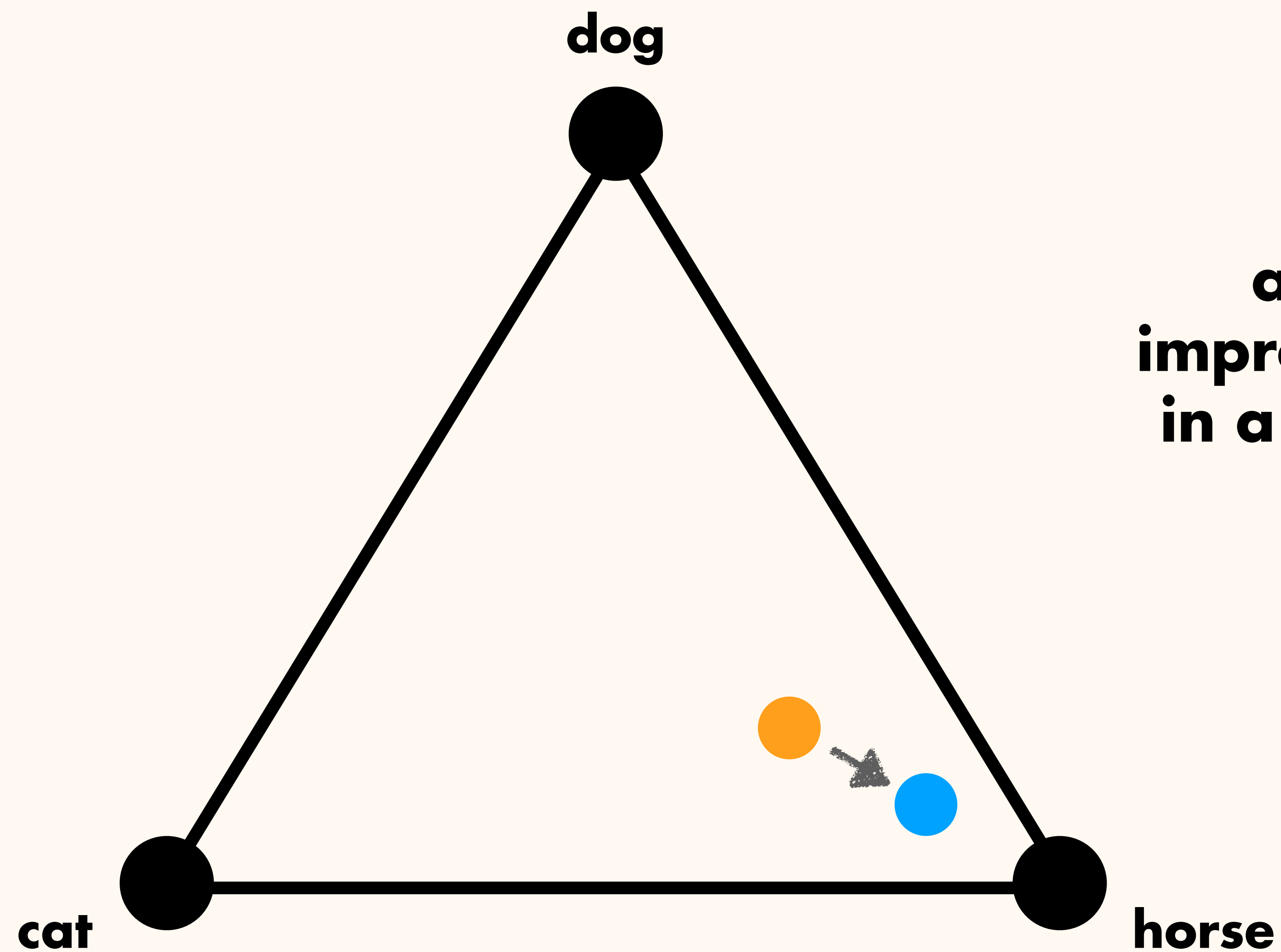
???



**The student
and the teacher
improve each other
in a virtuous cycle.**



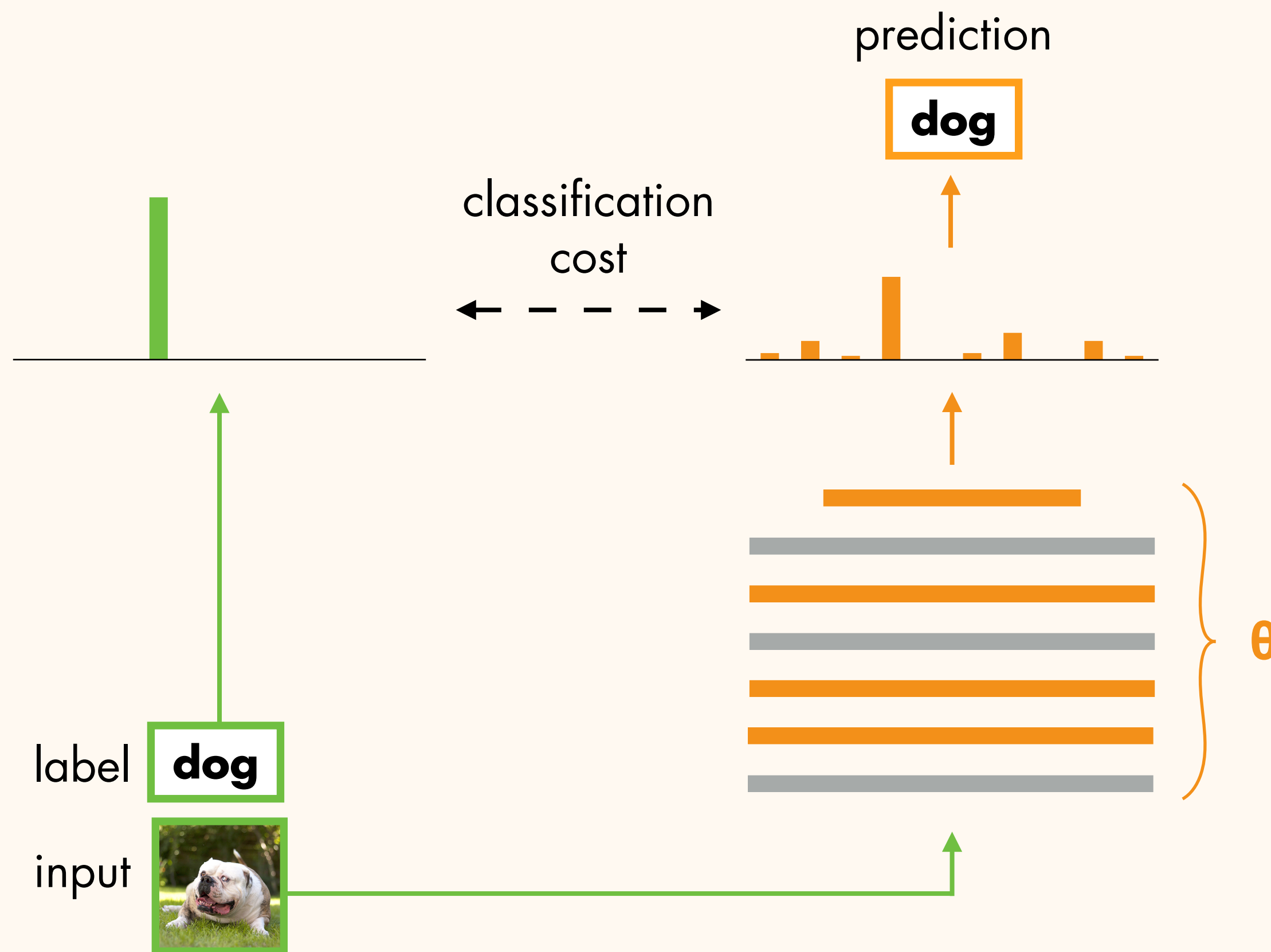
???



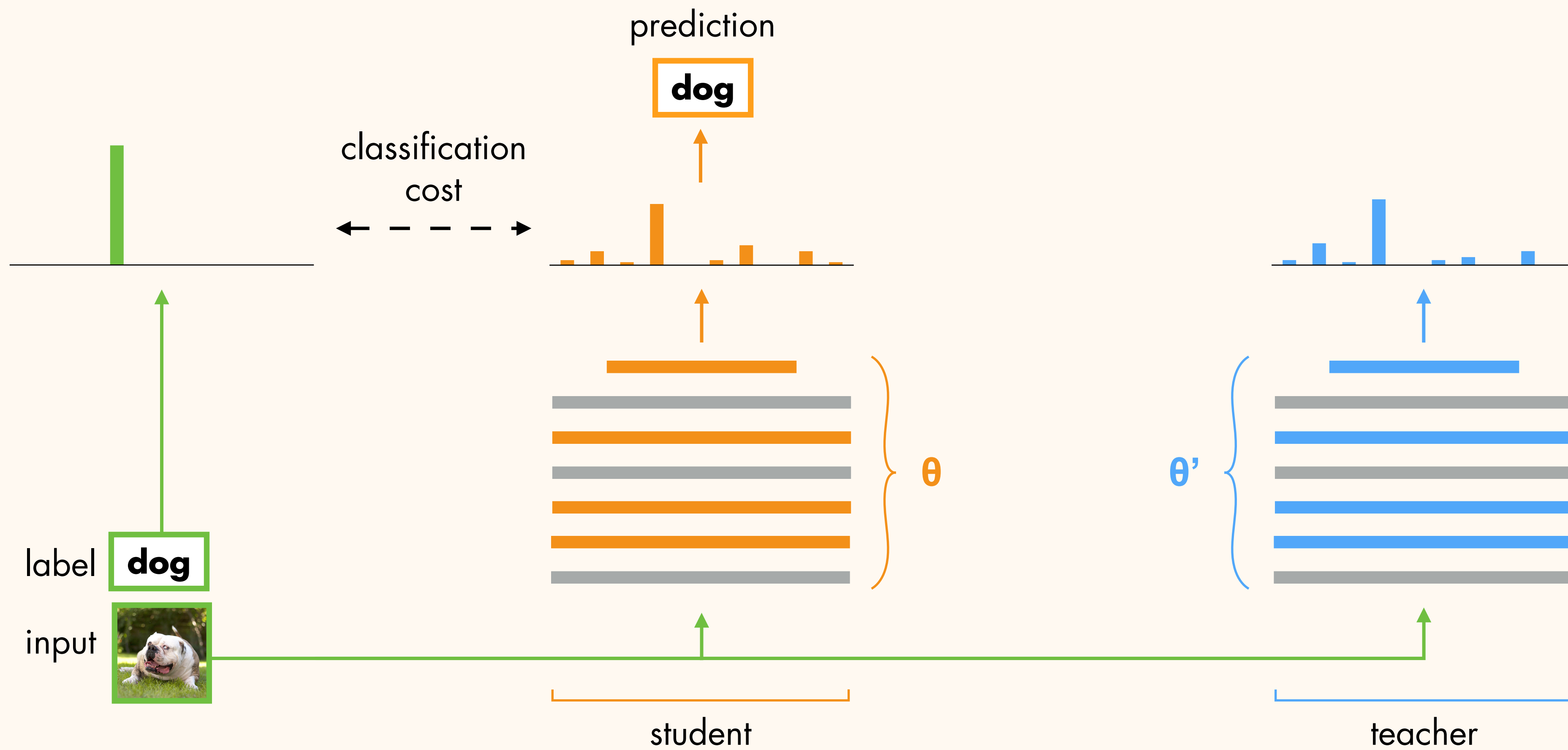
**The student
and the teacher
improve each other
in a virtuous cycle.**

HOW TO START USING THE MEAN TEACHER

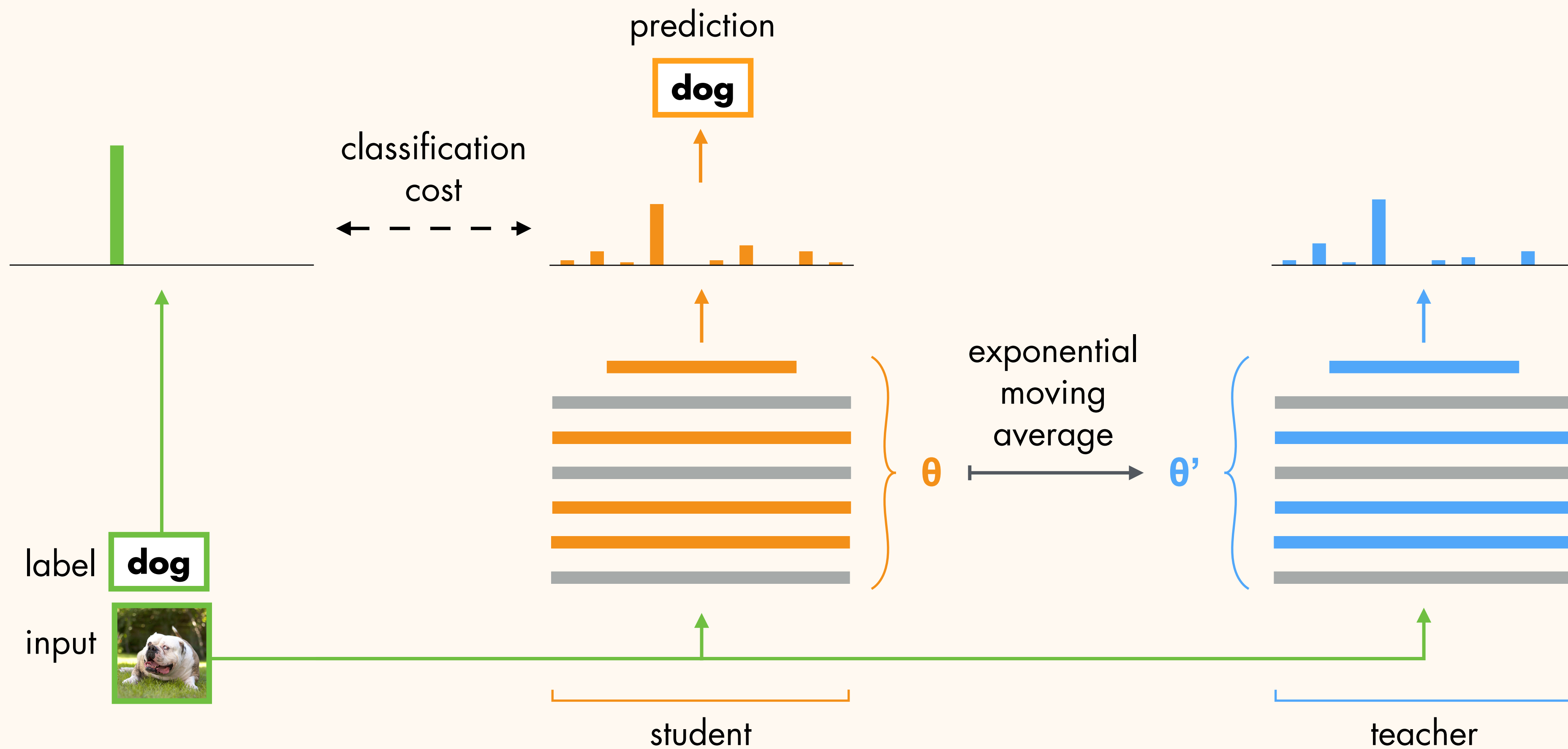
Take a supervised model.



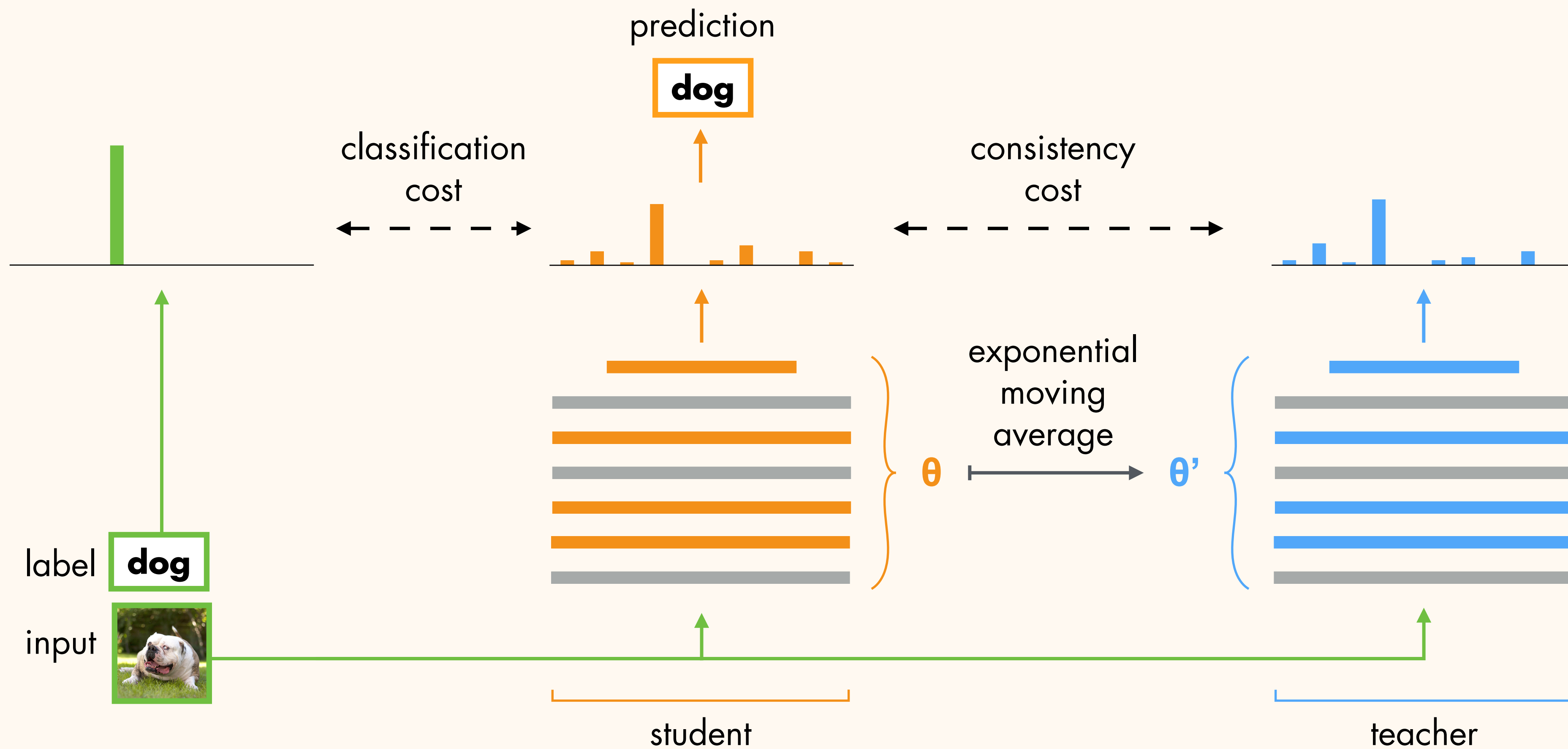
Make a copy of it.



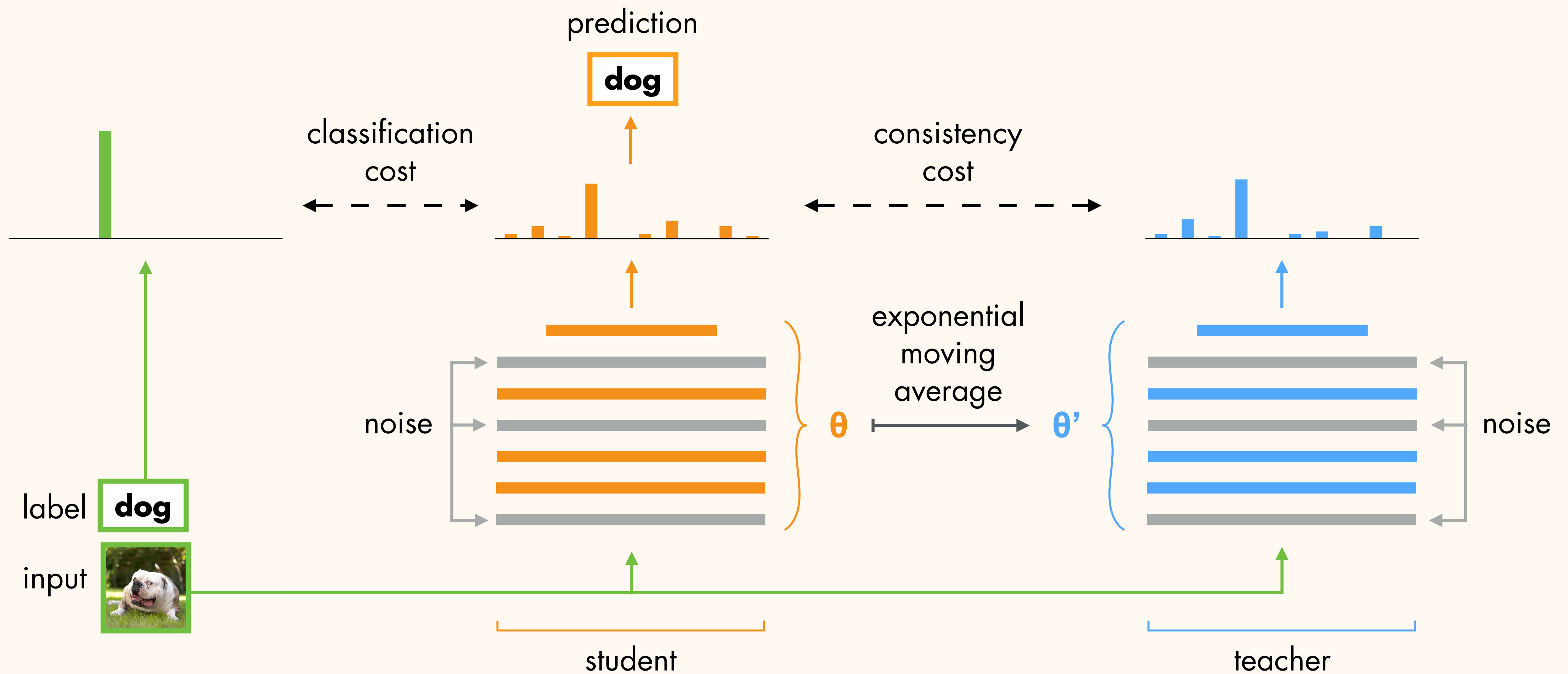
Update teacher weights after each training step.



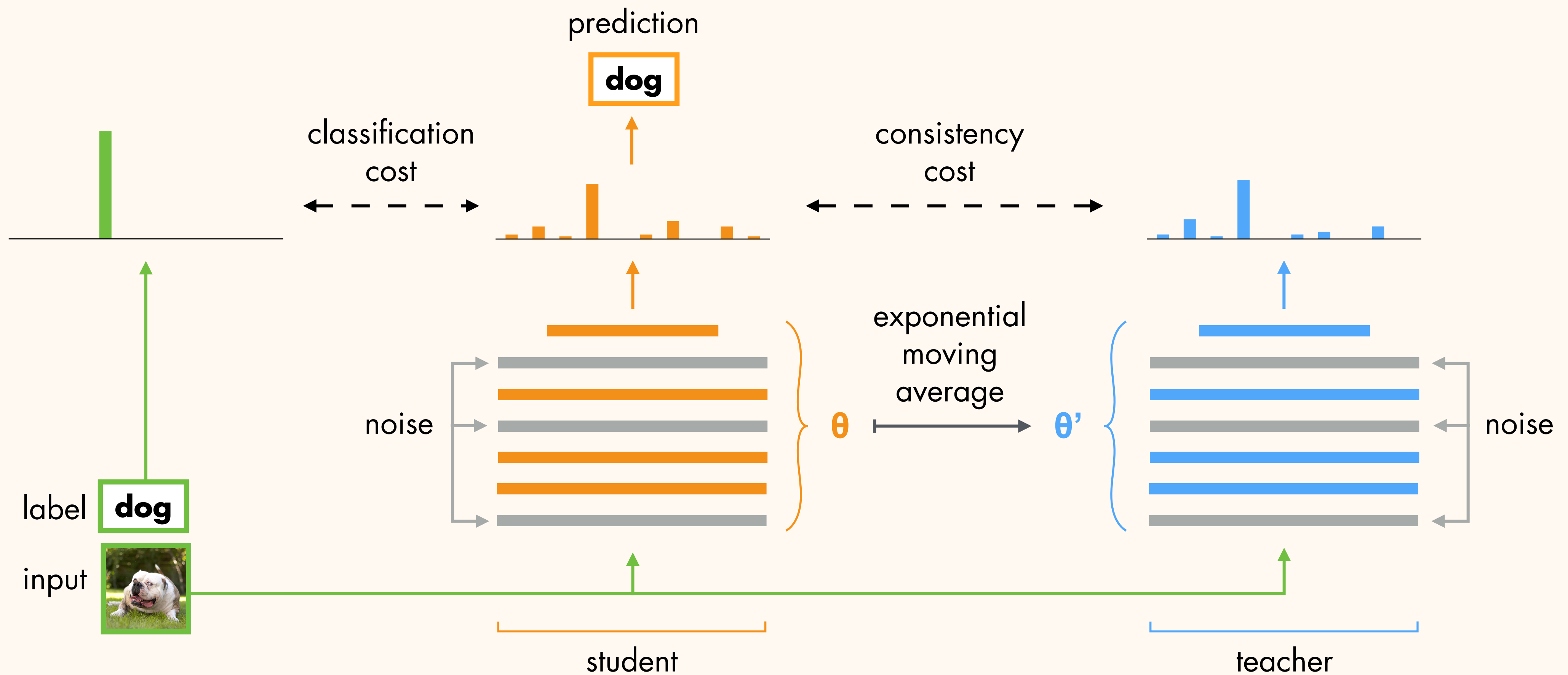
Add a cost between the two predictions.



Maybe add some noise.



Start using it for semi-supervised learning.



WHY TO USE THE MEAN TEACHER

WHY MEAN TEACHER

- 1. It is easy to add to your model.**
- 2. It can be adapted to different situations.**
- 3. It gives good results.**

WANT TO KNOW MORE?

PAPER

Mean Teachers Are Better Role Models

Weight-averaged consistency targets improve semi-supervised deep learning results

[arxiv:1703.01780](https://arxiv.org/abs/1703.01780)

POSTER

Today 06:30 – 10:30 PM
@ Pacific Ballroom #13

SOURCE CODE

<http://github.com/CuriousAI/mean-teacher>