

Siamese Network

Shusen Wang

Learning Pairwise Similarity Scores

Reference:

- Bromley et al. Signature verification using a Siamese time delay neural network. In *NIPS*. 1994.
- Koch, Zemel, & Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML*, 2015.

Training Set

Husky



⋮



Elephant



⋮



Tiger



⋮



Macaw



⋮



Car



⋮



Training Data

Positive Samples

(, , 1)

(, , 1)

(, , 1)

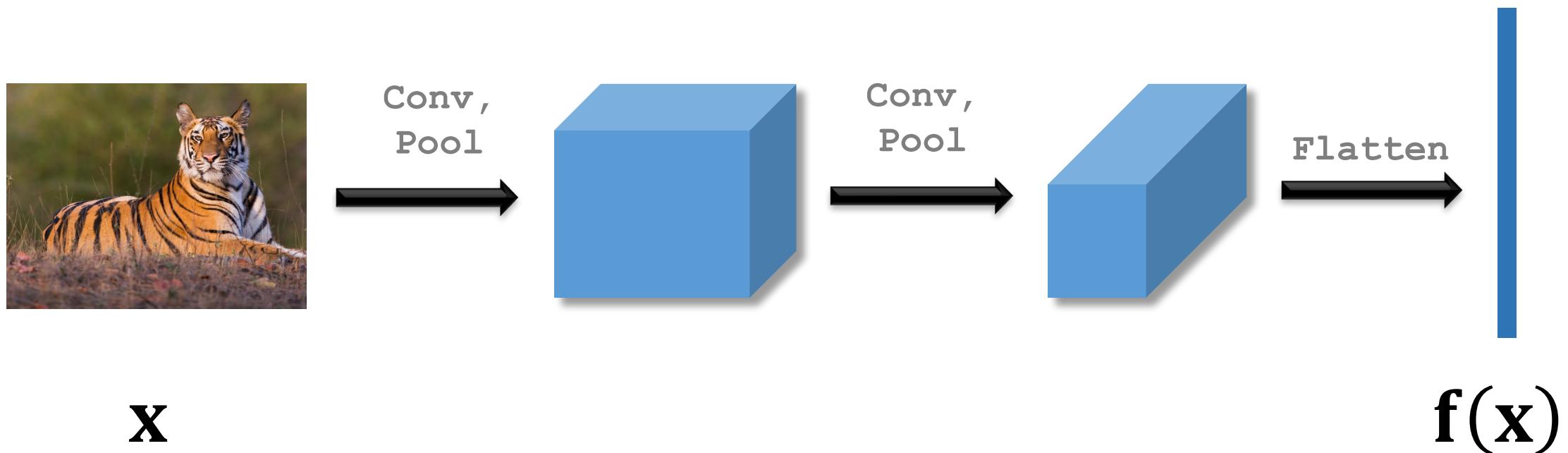
Negative Samples

(, , 0)

(, , 0)

(, , 0)

CNN for Feature Extraction



Training Siamese Network



\mathbf{x}_1



\mathbf{x}_2

Training Siamese Network



f



\mathbf{x}_1



f



\mathbf{x}_2

Training Siamese Network



\mathbf{x}_1

\mathbf{f}



$$\mathbf{h}_1 = \mathbf{f}(\mathbf{x}_1)$$



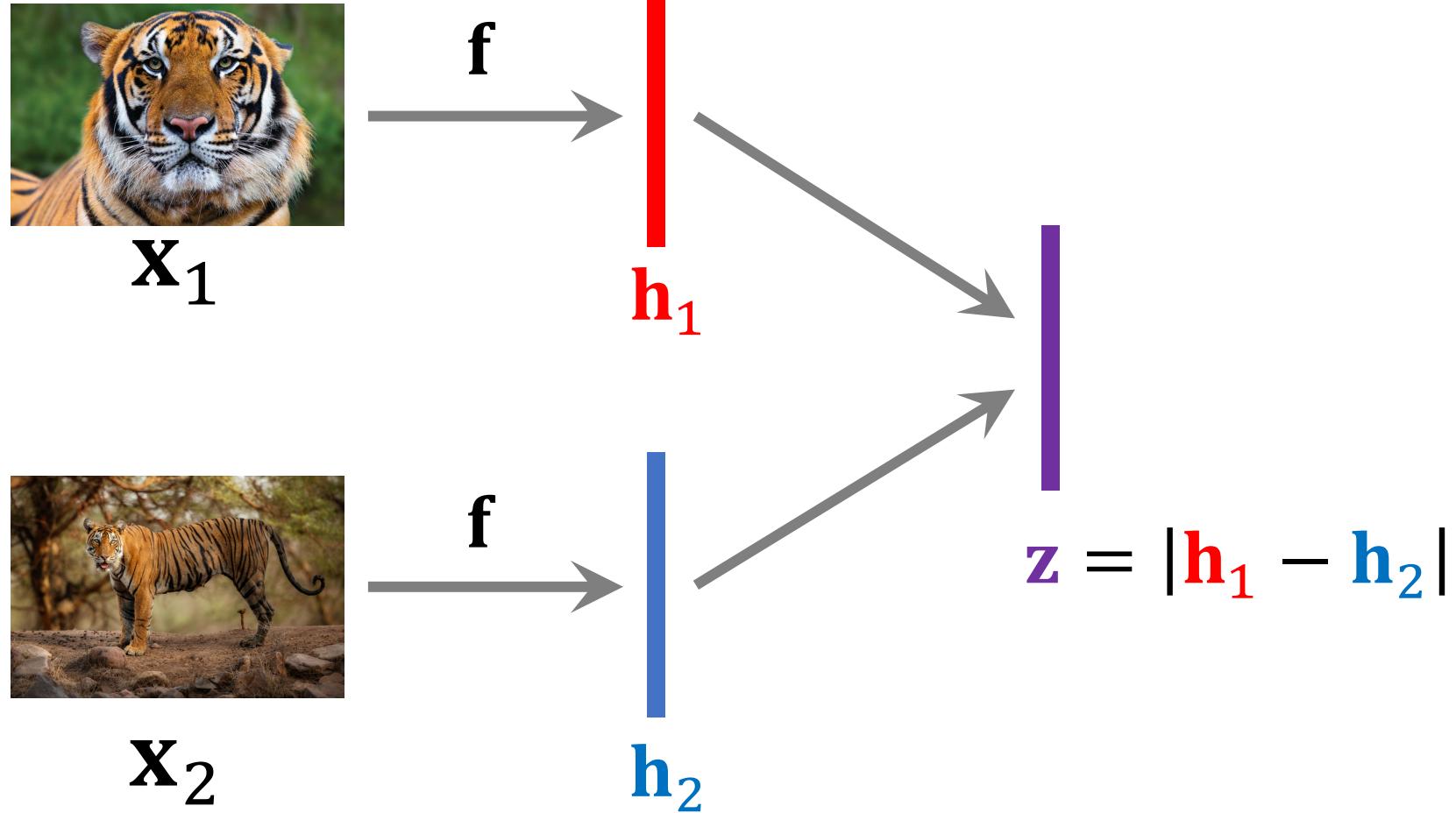
\mathbf{x}_2

\mathbf{f}

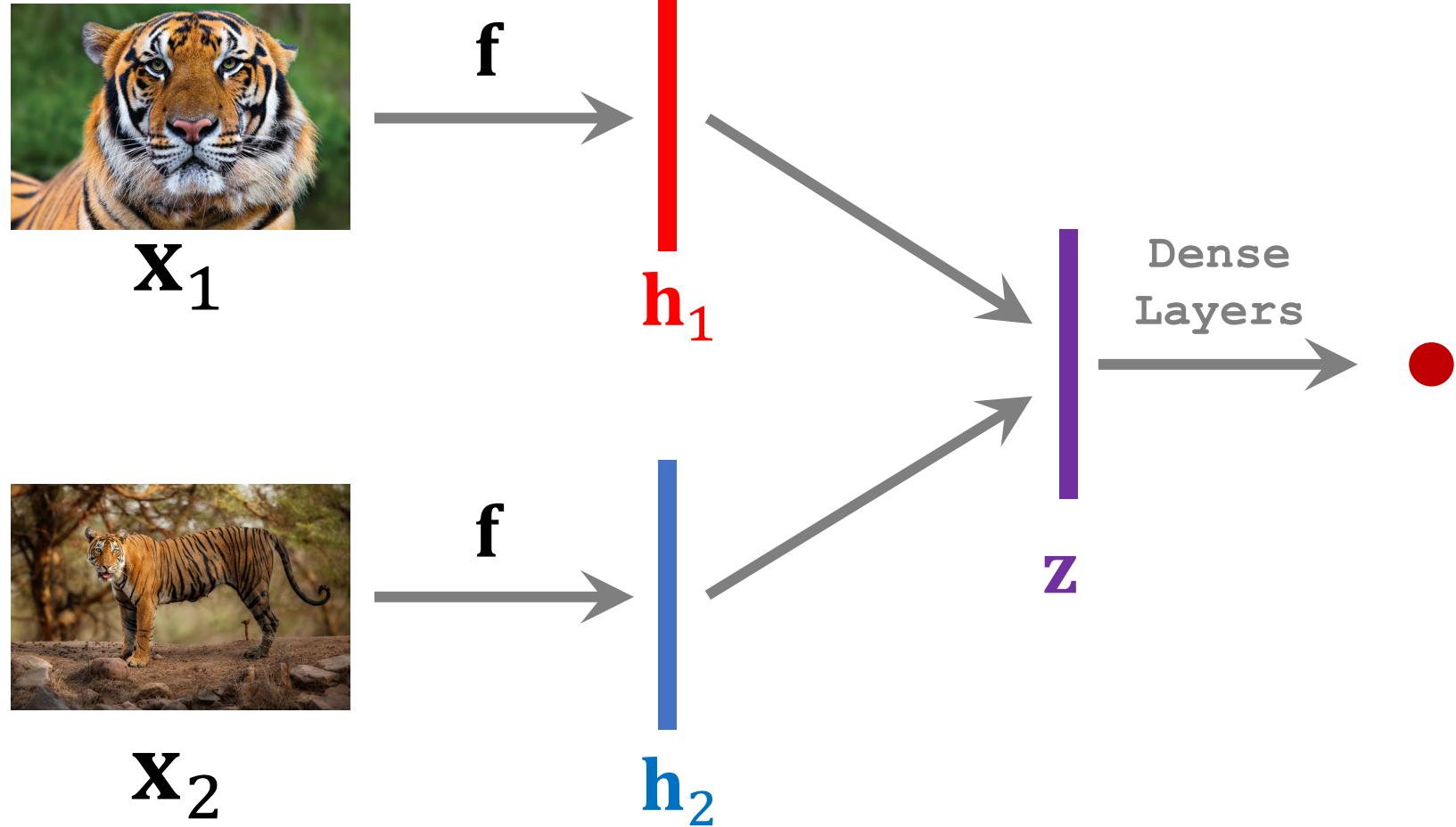


$$\mathbf{h}_2 = \mathbf{f}(\mathbf{x}_2)$$

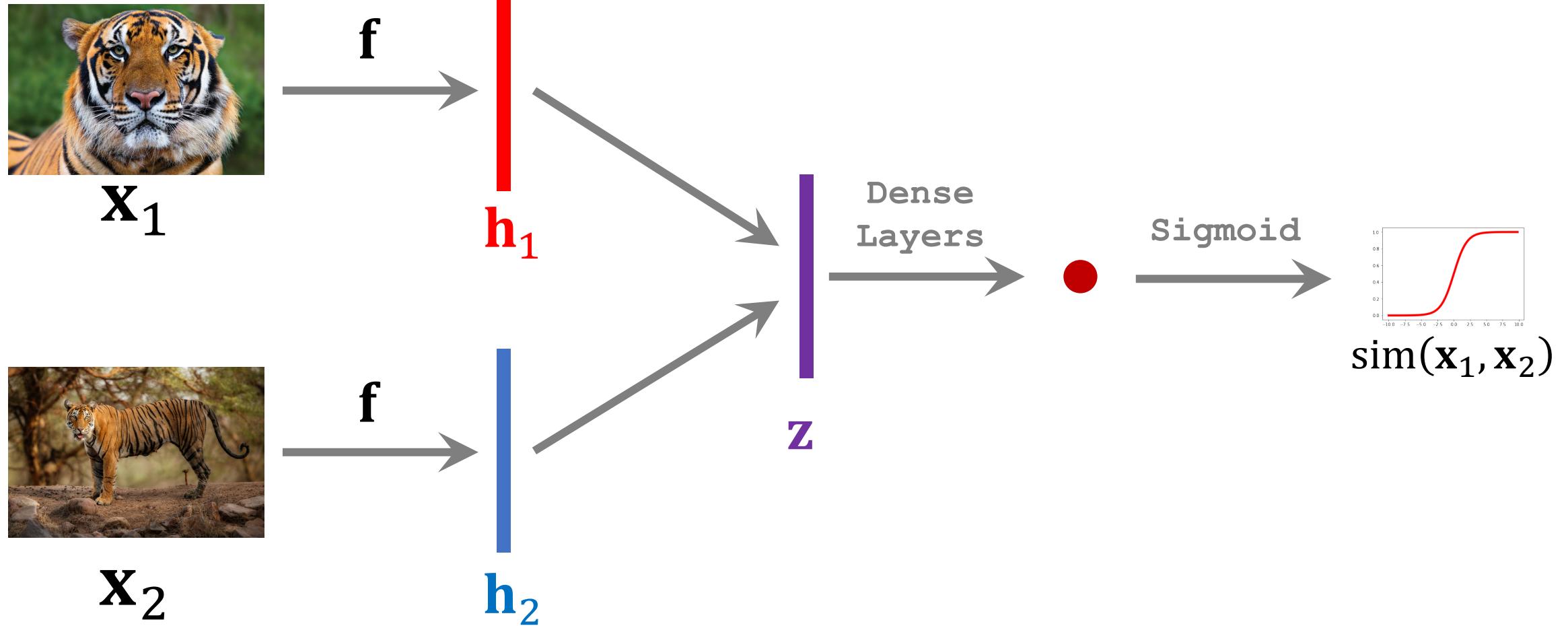
Training Siamese Network



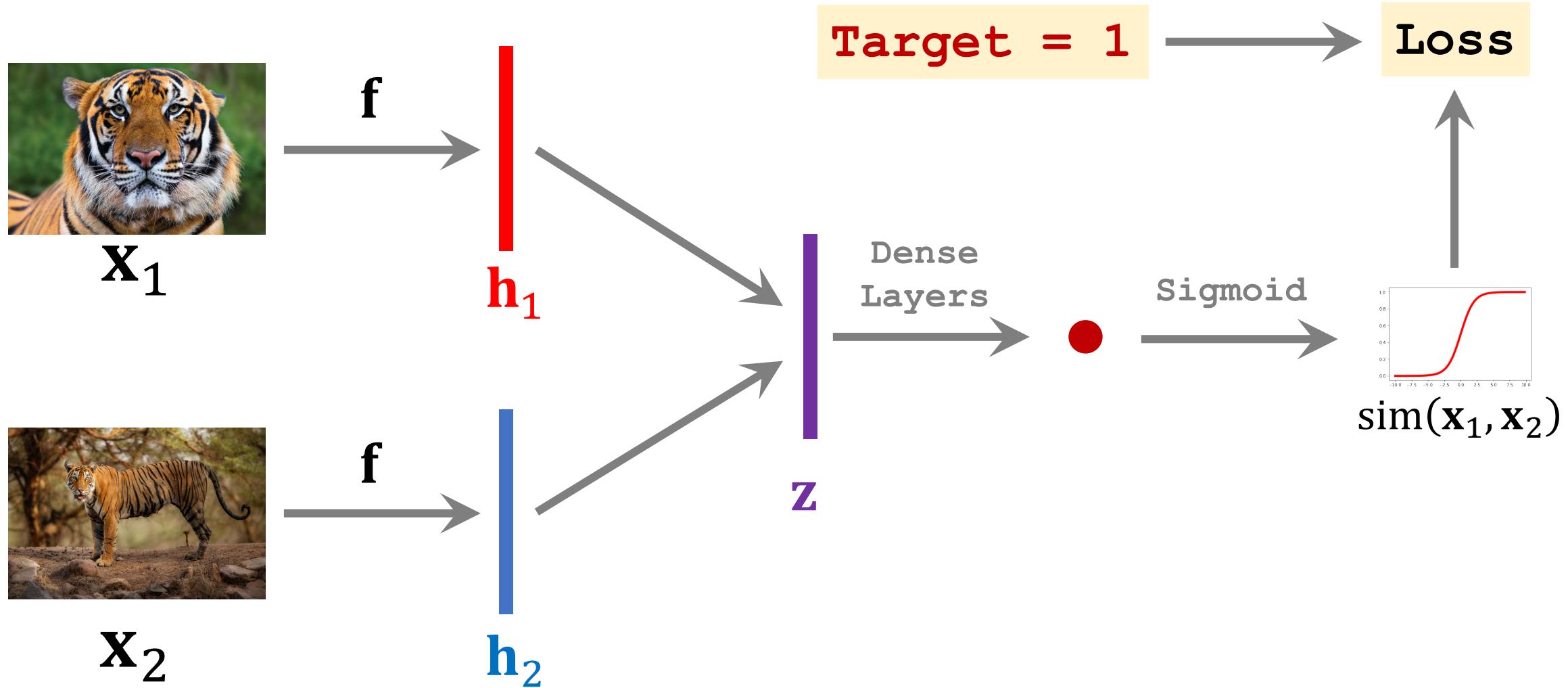
Training Siamese Network



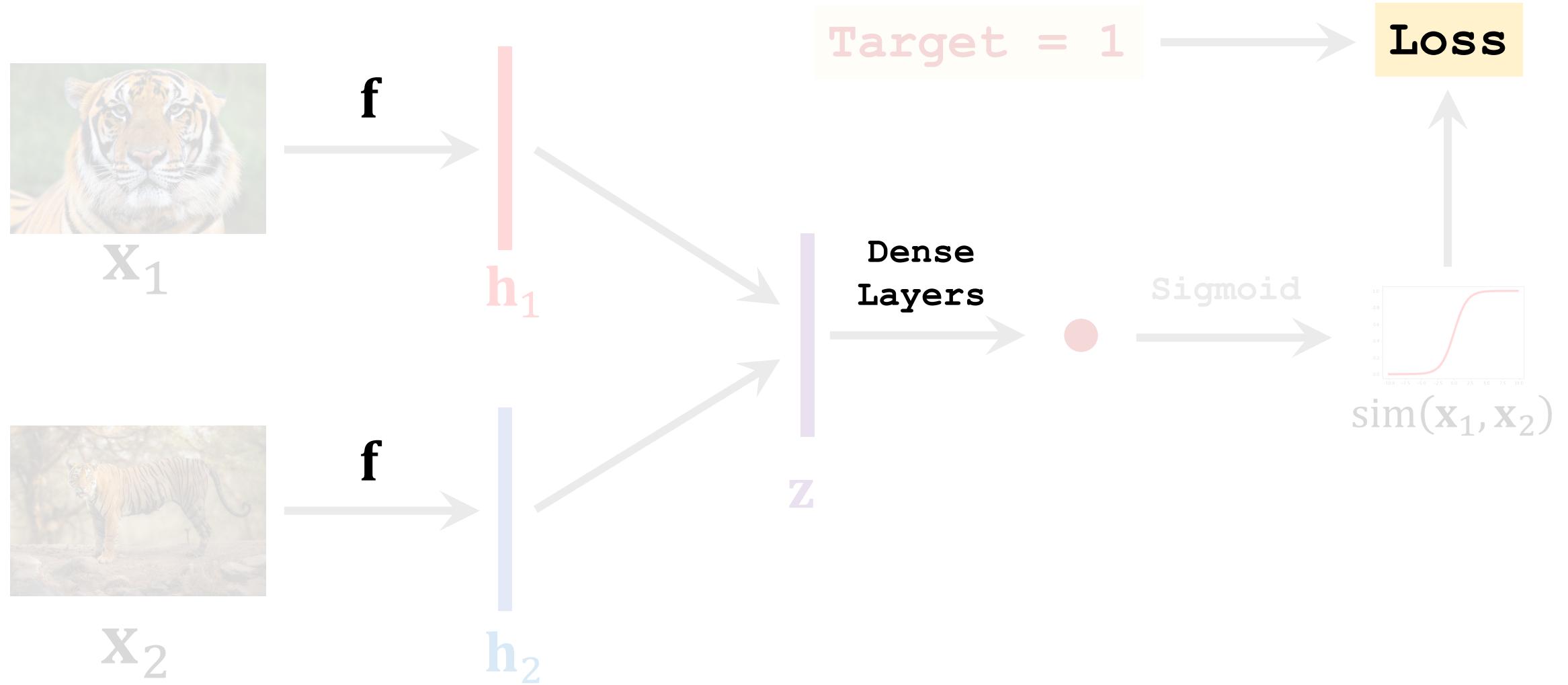
Training Siamese Network



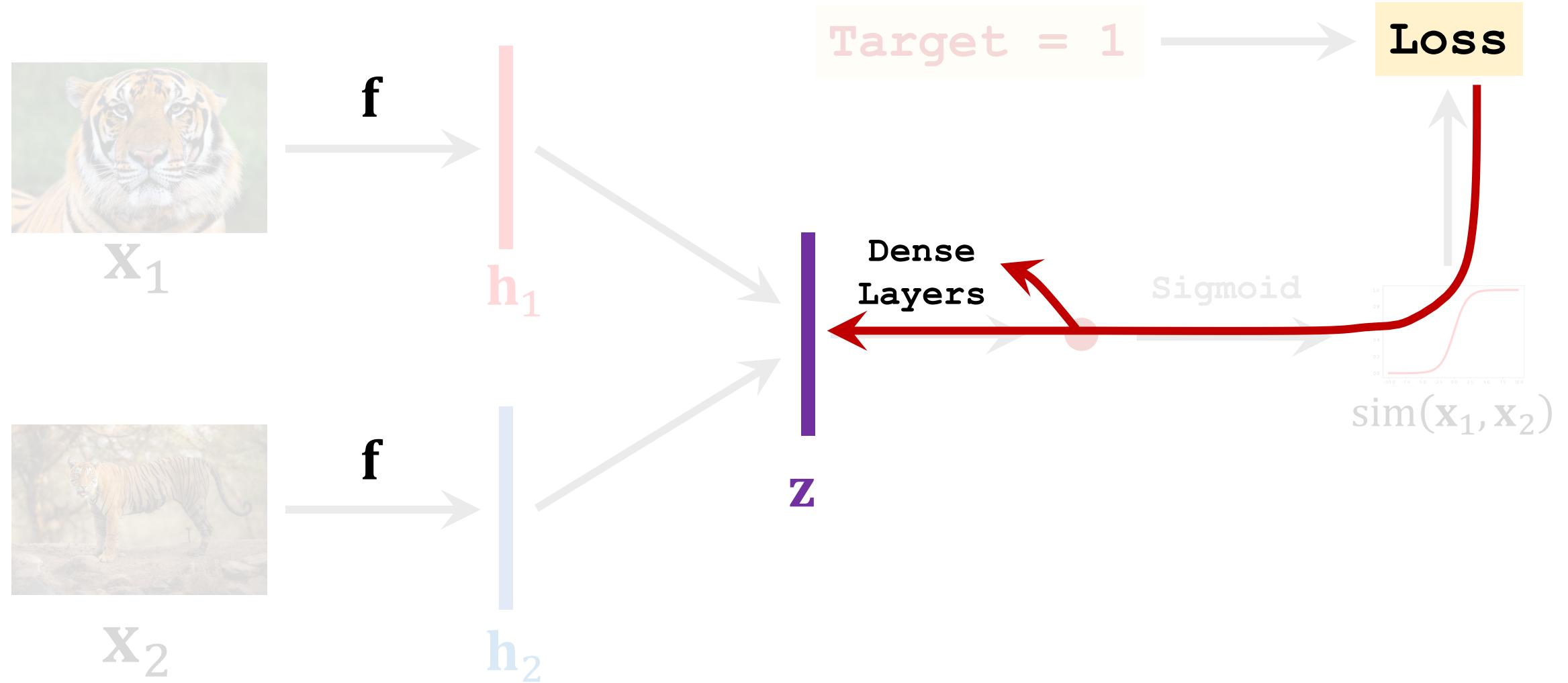
Training Siamese Network



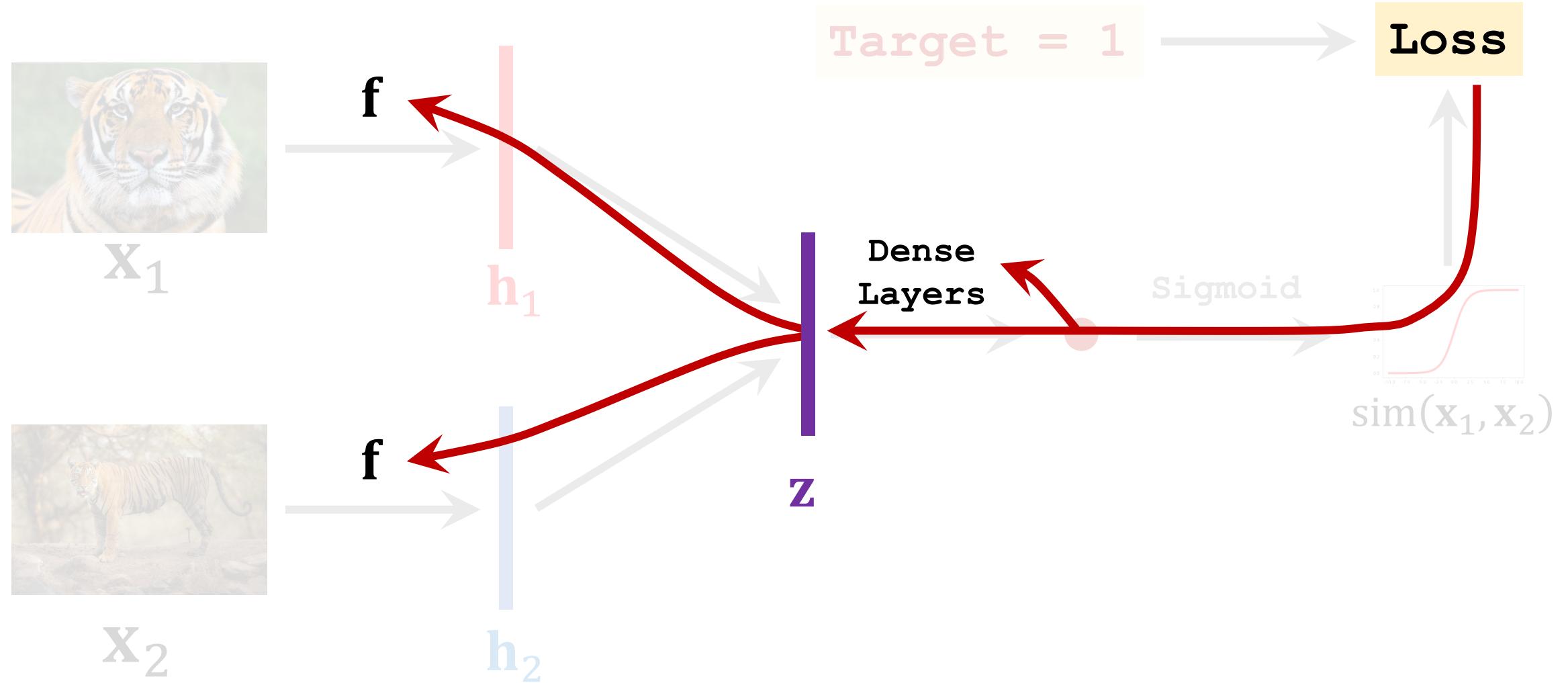
Training Siamese Network



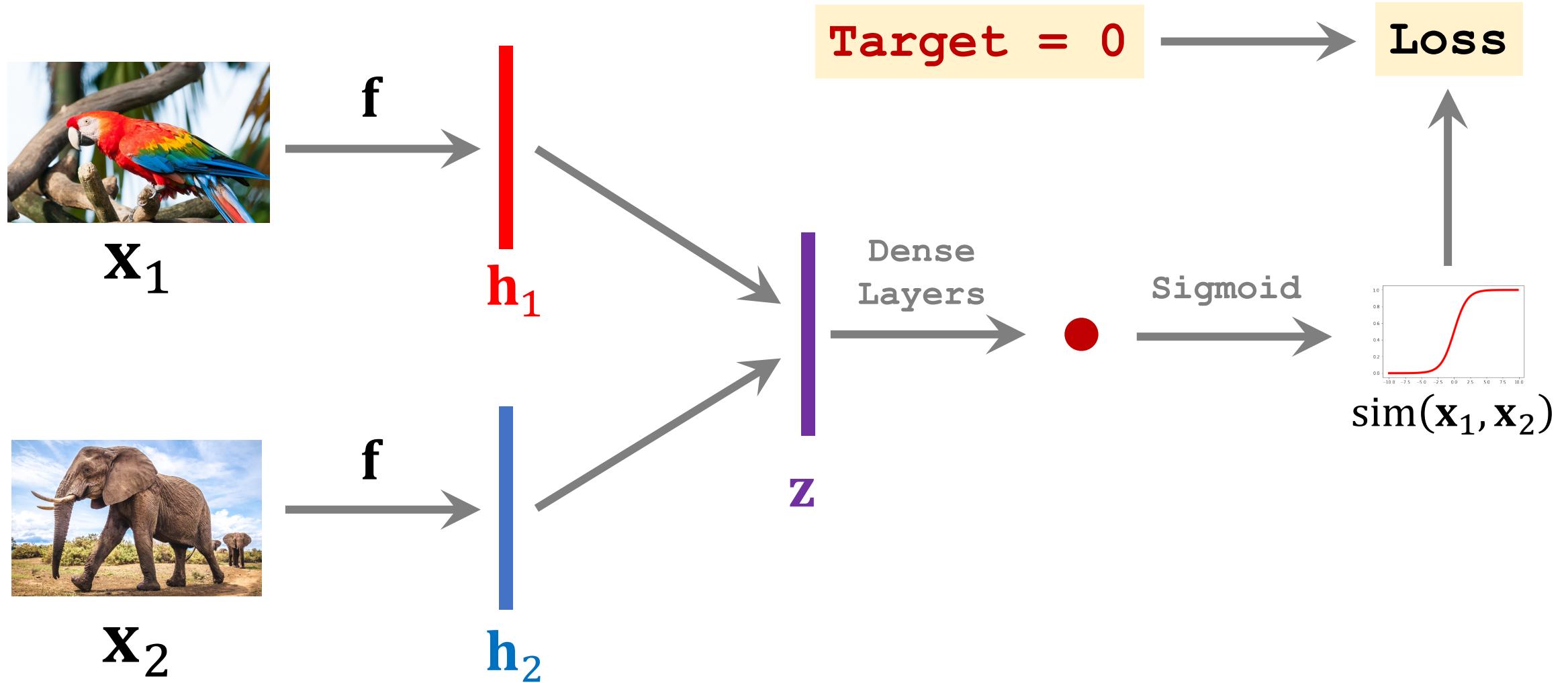
Training Siamese Network



Training Siamese Network



Training Siamese Network



One-Shot Prediction

- 6-way 1-shot prediction: support set has 6 test classes; each class has 1 sample.
- The training data (for the Siamese network) does not contain the 6 classes.

Support Set:

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



Support Set:

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



sim = 0.2

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



sim = 0.2

sim = 0.9

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



sim = 0.2

sim = 0.9

sim = 0.7

sim = 0.5

sim = 0.3

sim = 0.4

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



sim = 0.2



sim = 0.9



sim = 0.7



sim = 0.5



sim = 0.3



sim = 0.4



Triplet Loss

Reference:

- Schroff, Kalenichenko, & Philbin. [Facenet: A unified embedding for face recognition and clustering](#). In *CVPR*, 2015.

Data for Training Siamese Network

Training Set

Elephant



Tiger



Macaw



Car



Data for Training Siamese Network

Training Set

Elephant



Tiger



Macaw



Car



anchor

Data for Training Siamese Network

Training Set



Elephant



Tiger



anchor

Macaw



Car



Data for Training Siamese Network

Training Set



Elephant



Tiger



Macaw



Car



Data for Training Siamese Network



Training Set

Elephant



Tiger



Macaw



Car



Data for Training Siamese Network



Training Set

Elephant



Negative

Tiger



Macaw



Car



Data for Training Siamese Network



Training Set

Elephant



Tiger



Macaw



Car



Triplet Loss



\mathbf{x}^+
(positive)



\mathbf{x}^a
(anchor)



\mathbf{x}^-
(negative)

Triplet Loss



x^+
(positive)

$$f \longrightarrow$$



x^a
(anchor)

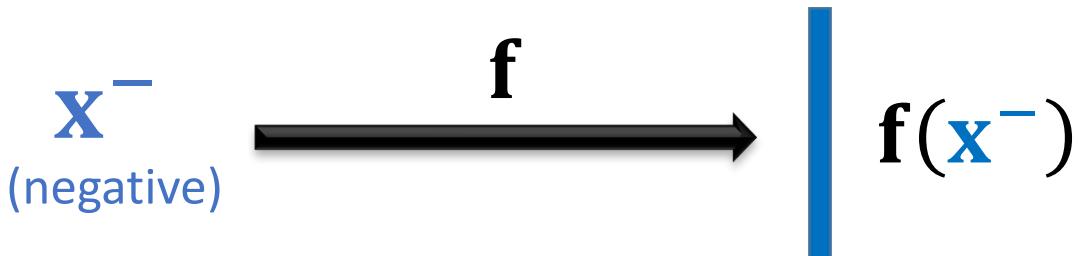
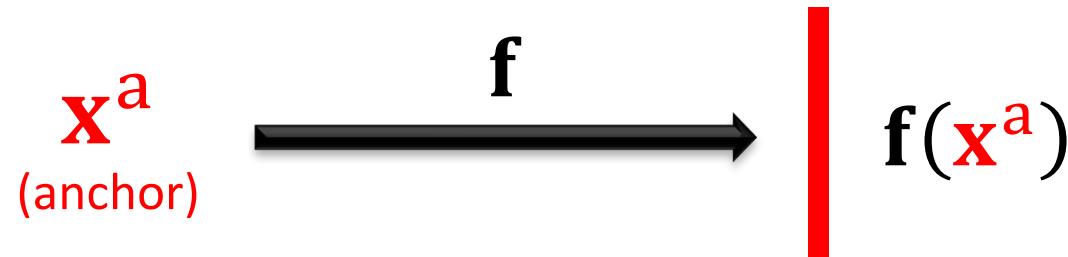
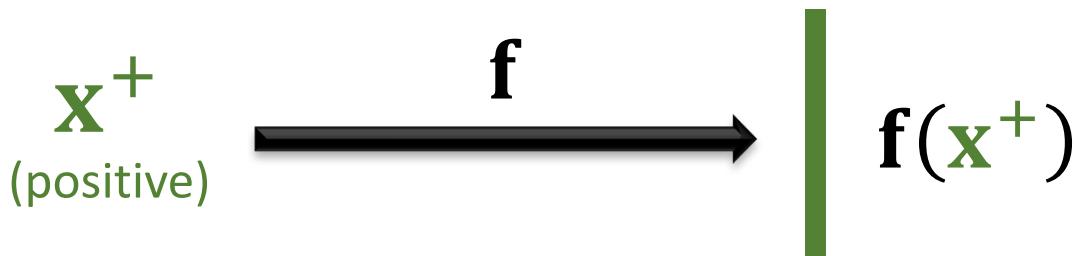
$$f \longrightarrow$$



x^-
(negative)

$$f \longrightarrow$$

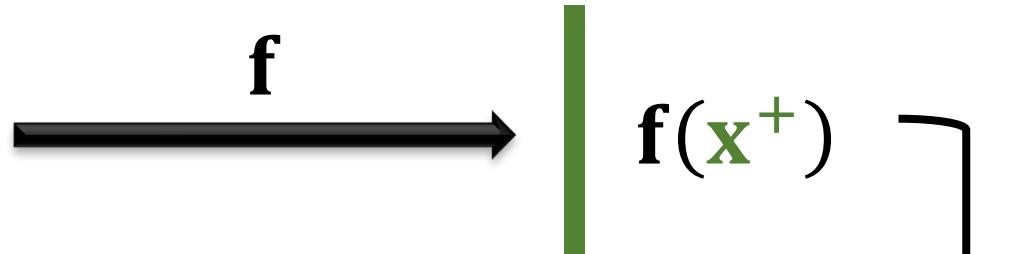
Triplet Loss



Triplet Loss



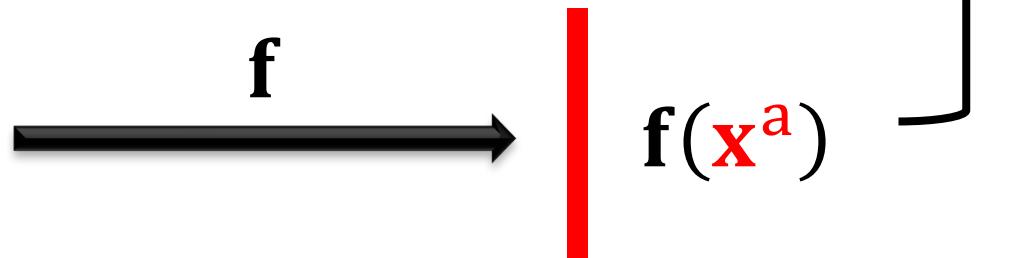
\mathbf{x}^+
(positive)



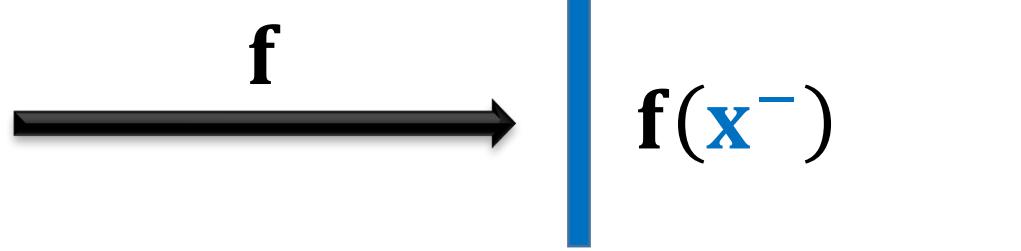
$$d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$$



\mathbf{x}^a
(anchor)



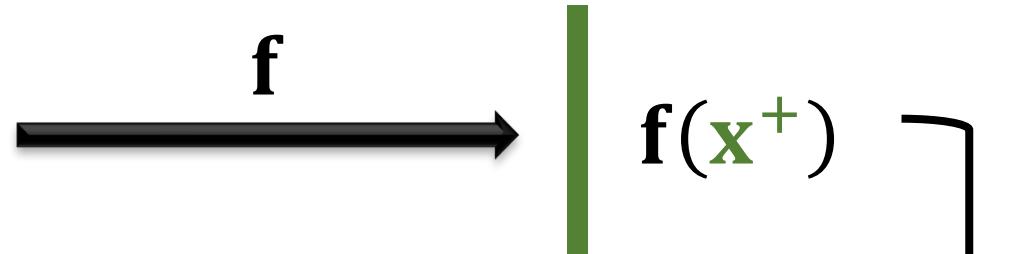
\mathbf{x}^-
(negative)



Triplet Loss



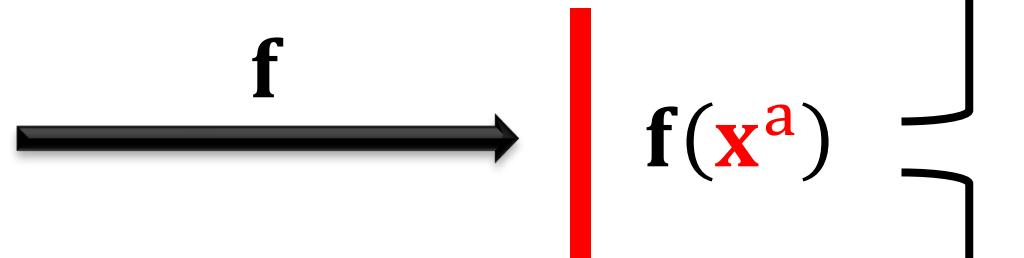
\mathbf{x}^+
(positive)



$$d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$$



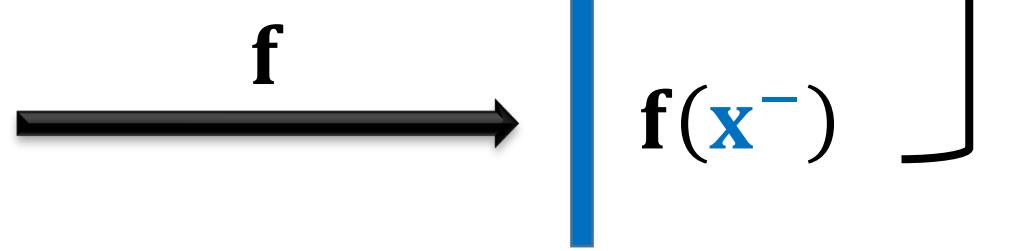
\mathbf{x}^a
(anchor)



$$d^- = \|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\|_2^2$$



\mathbf{x}^-
(negative)



Triplet Loss

Feature Space



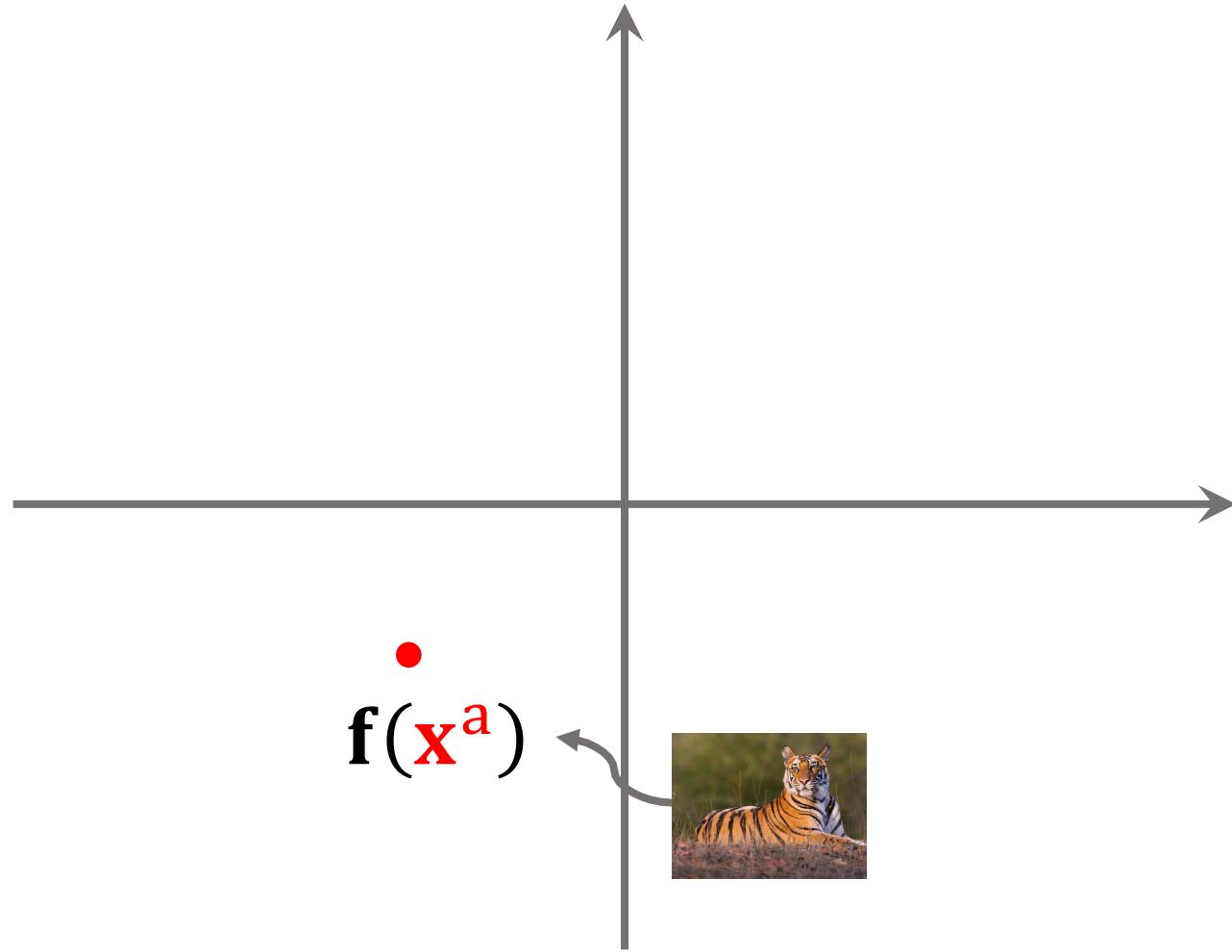
x^+
(positive)



x^a
(anchor)



x^-
(negative)



Triplet Loss



x^+
(positive)

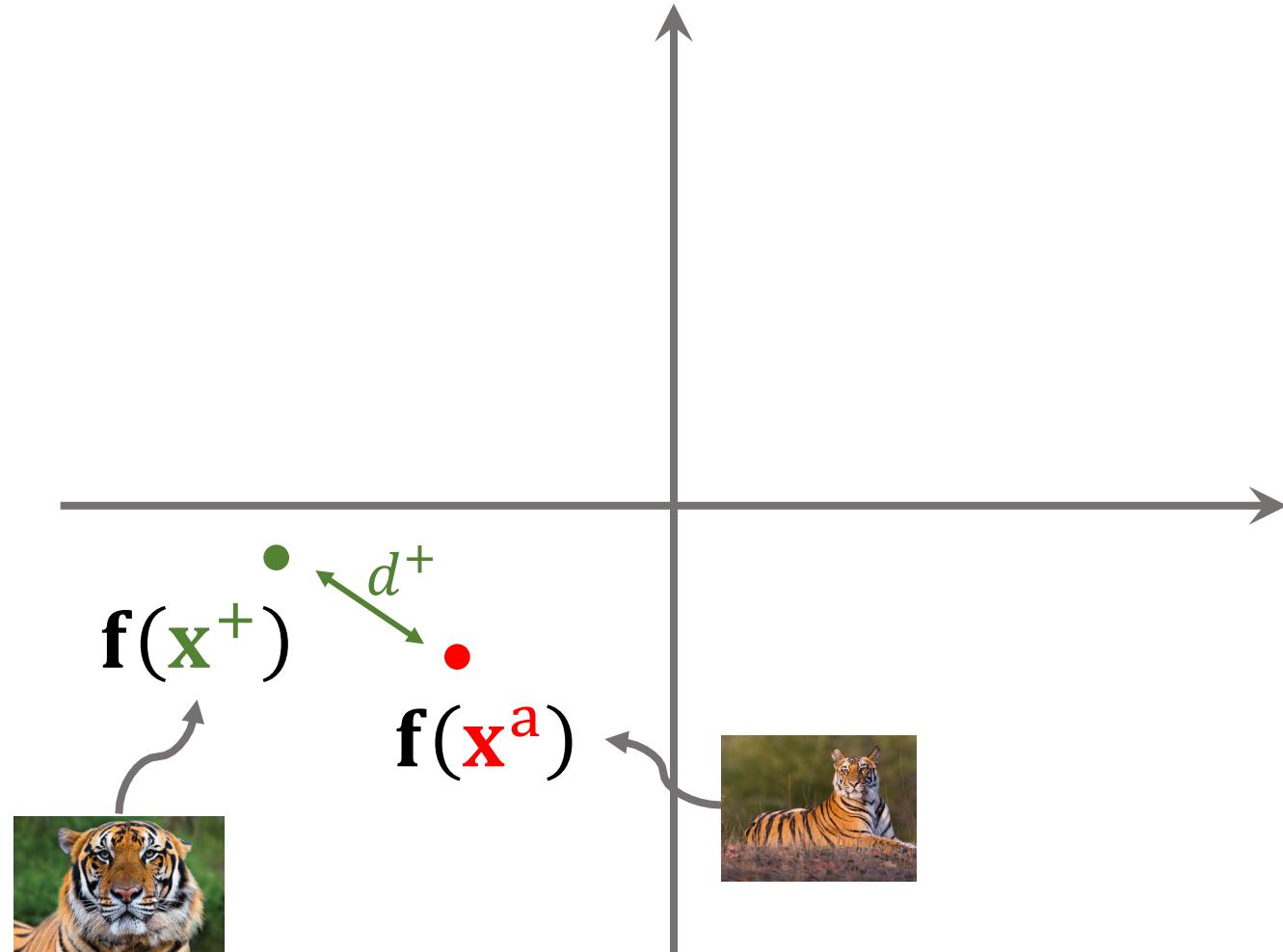


x^a
(anchor)



x^-
(negative)

Feature Space



Triplet Loss



\mathbf{x}^+
(positive)

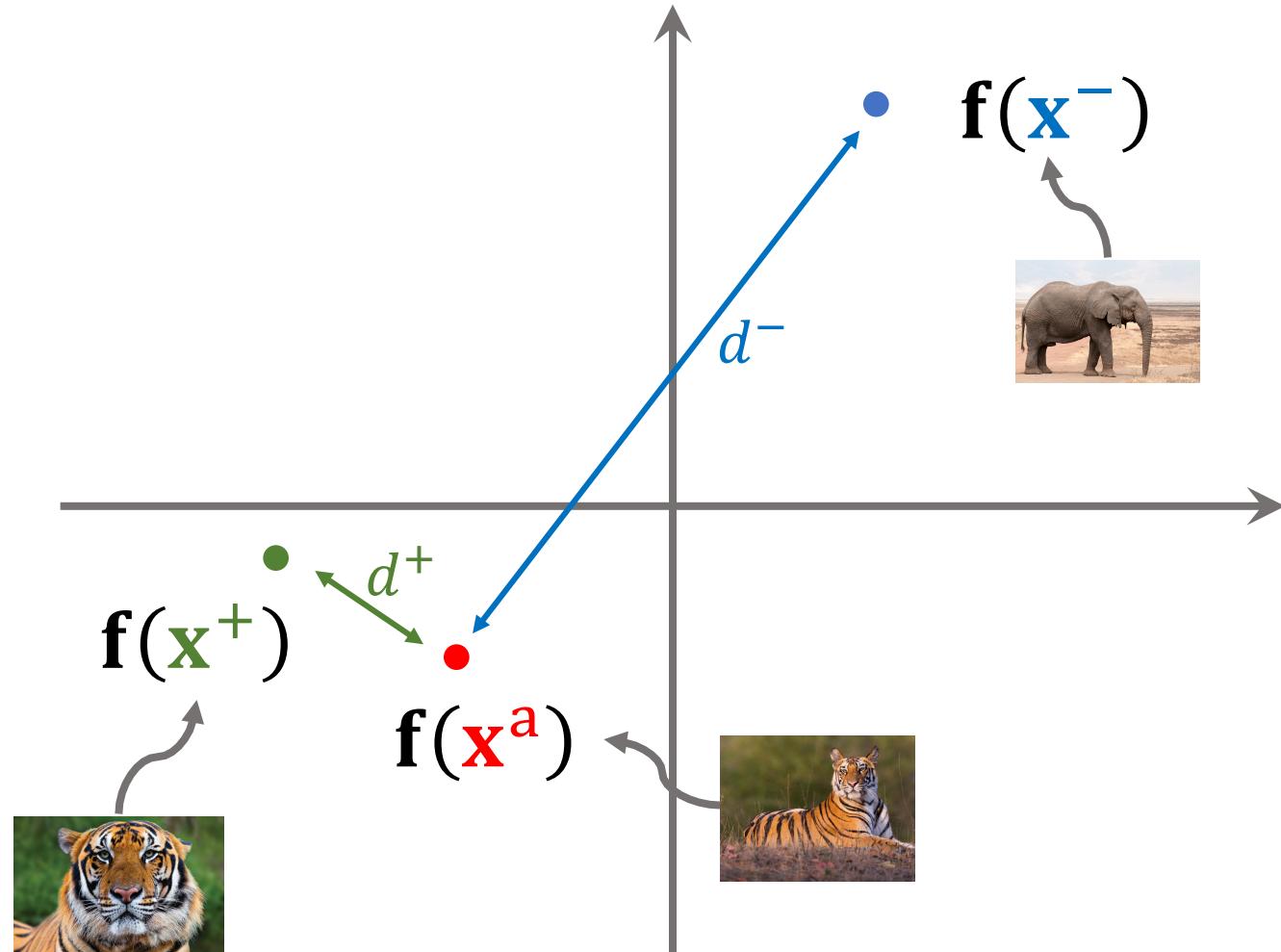


\mathbf{x}^a
(anchor)



\mathbf{x}^-
(negative)

Feature Space



Triplet Loss



\mathbf{x}^+
(positive)



\mathbf{x}^a
(anchor)



\mathbf{x}^-
(negative)

- Encourage $d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$ to be small.
- Encourage $d^- = \|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\|_2^2$ to be big.

Triplet Loss



\mathbf{x}^+
(positive)



\mathbf{x}^a
(anchor)



\mathbf{x}^-
(negative)

- Encourage $d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$ to be small.
- Encourage $d^- = \|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\|_2^2$ to be big.
- If $d^- \geq d^+ + \alpha$, then no loss. ($\alpha > 0$ is margin.)
- Otherwise, the loss is $d^+ + \alpha - d^-$.

Triplet Loss



\mathbf{x}^+
(positive)



\mathbf{x}^a
(anchor)



\mathbf{x}^-
(negative)

- Encourage $d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$ to be small.
- Encourage $d^- = \|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\|_2^2$ to be big.
- If $d^- \geq d^+ + \alpha$, then no loss. ($\alpha > 0$ is margin.)
- Otherwise, the loss is $d^+ + \alpha - d^-$.
- $\text{Loss}(\mathbf{x}^a, \mathbf{x}^+, \mathbf{x}^-) = \max\{0, d^+ + \alpha - d^-\}$.
- Update the CNN (function \mathbf{f}) to decrease the loss.

One-Shot Prediction

Query:



Support Set:

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



dist = 231

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



dist = 231

dist = 19

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



dist = 231

dist = 19

dist = 138

dist = 76

dist = 122

dist = 94

Fox



Squirrel



Rabbit



Hamster



Otter



Beaver



One-Shot Prediction

Query:



dist = 231



dist = 19



dist = 138



dist = 76



dist = 122



dist = 94

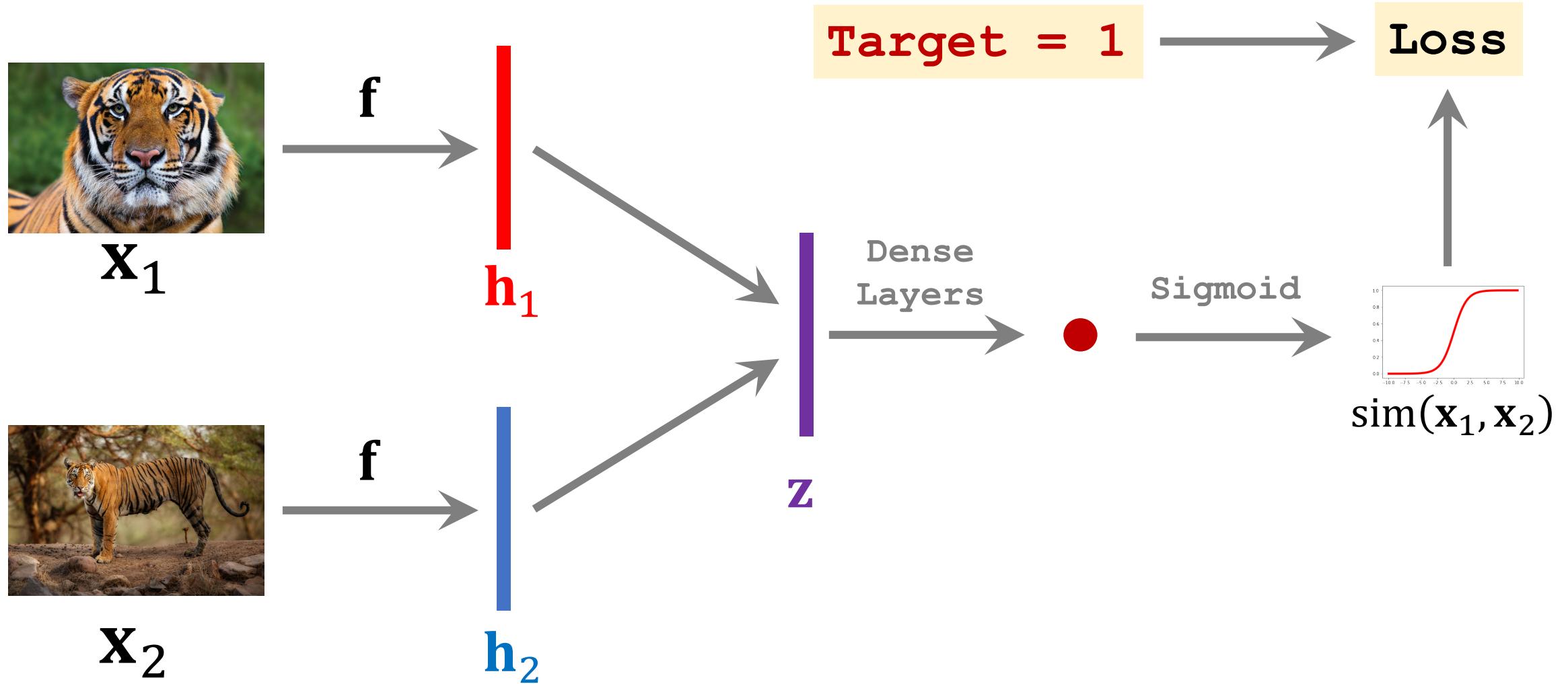


Summary

Basic Idea of Few-Shot Learning

- Train a **Siamese network** on large-scale training set.
- Given a **support set** of k -way n -shot.
 - k -way means k classes.
 - n -shot means every class has n samples.
 - The training set does not contain the k classes.
- Given a **query**, predict its class.
 - Use the Siamese network to compute similarity or distance.

Siamese Network for Pairwise Similarity

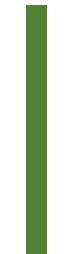


Siamese Network with Triplet Loss



\mathbf{x}^+
(positive)

$$\mathbf{x}^+ \xrightarrow{\mathbf{f}}$$



$\mathbf{f}(\mathbf{x}^+)$



\mathbf{x}^a
(anchor)

$$\mathbf{x}^a \xrightarrow{\mathbf{f}}$$



$\mathbf{f}(\mathbf{x}^a)$

$$d^+ = \|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\|_2^2$$



\mathbf{x}^-
(negative)

$$\mathbf{x}^- \xrightarrow{\mathbf{f}}$$



$\mathbf{f}(\mathbf{x}^-)$

$$d^- = \|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\|_2^2$$

Thank you!