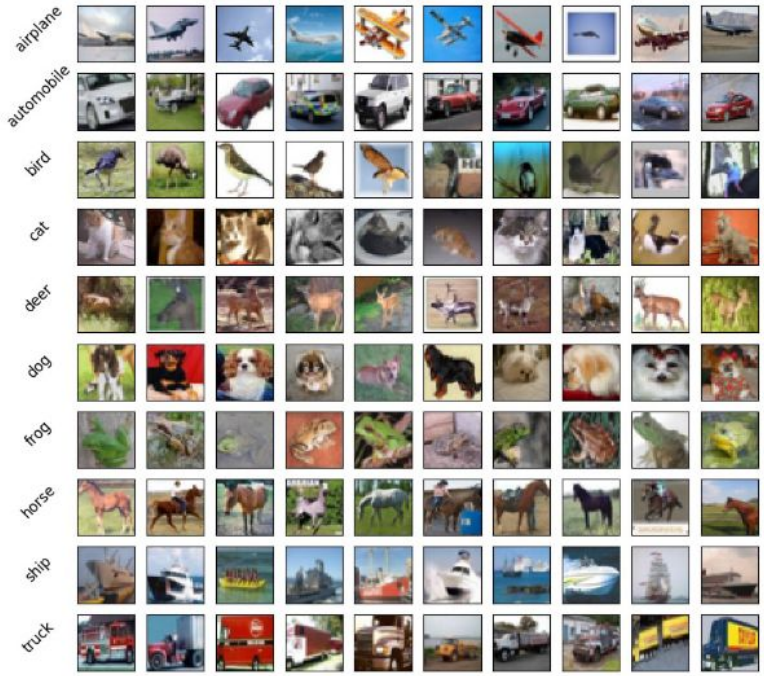
Optimization

Artificial Intelligence
University of the Philippines Diliman
2022

What is optimization in the context of Machine Learning?



Input : $p(\boldsymbol{x})$

f(θ)

ML Model

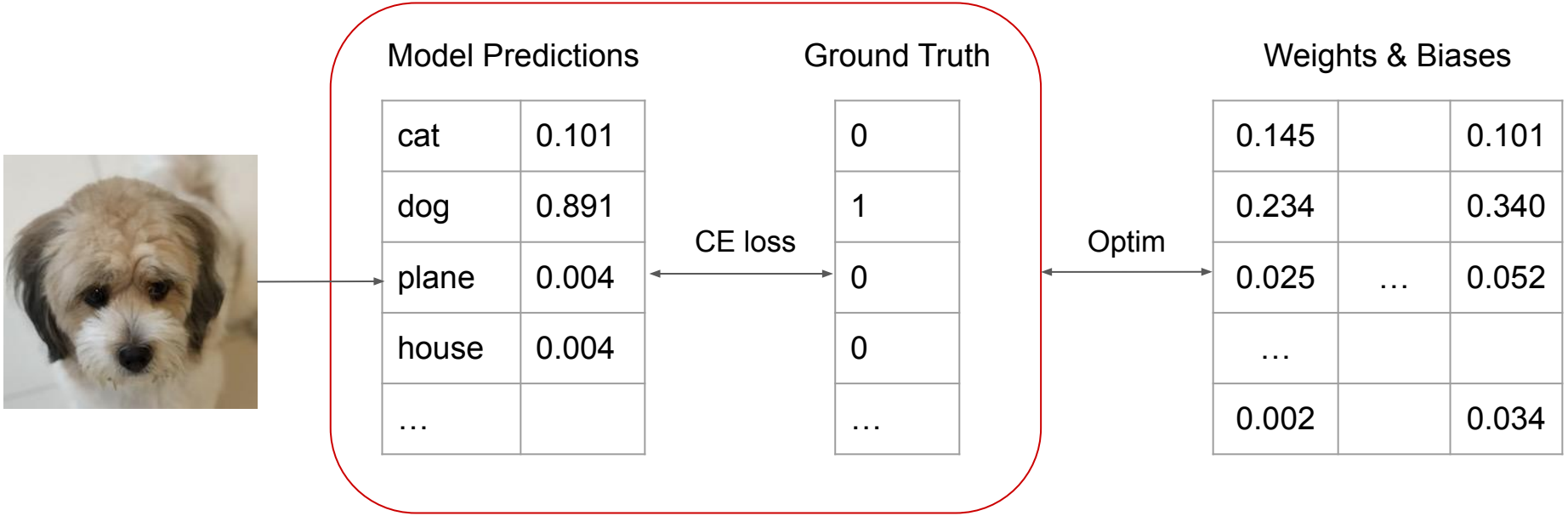
Output: $p(\boldsymbol{y}|\boldsymbol{x})$

Parameters : $p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$

Source

Loss function and optimization

- Loss function: What to minimize?
- Optimizer: How to minimize?



Common DL Optimizers

For full list, *see* PyTorch

	Pros	Cons
SGD	Basic algorithm Less memory requirements	Slow compute time Constant learning rate Potentially noisy
RMSProp	Faster and less tuning vs SGD	Initial learning rate is manual
AdaGrad	Good for sparse data Adaptive learning rate	Prone to vanishing gradients
AdaDelta	Pros of AdaGrad Prevents vanishing gradients	Computationally expensive
Adam	Less tuning Low memory requirement, etc	May not converge to optimum solution in some scenarios