SEP 22-26

AIM for Scale
AI Weather Training Program

AIM FOR SCALE | WEATHER PACKAGE

AI Overview

Souhaib Ben Taieb

What is Artificial Intelligence?

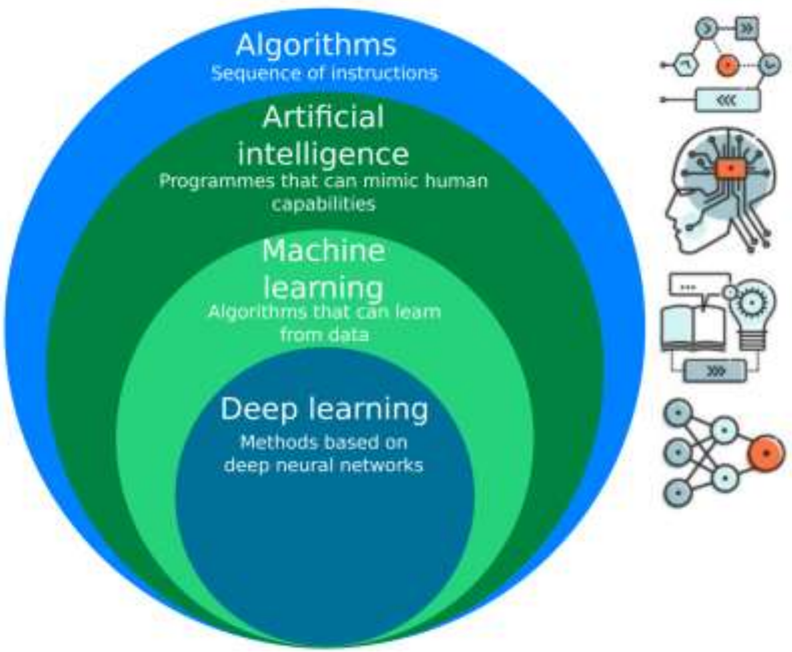
The science of making computers do things that human beings can do

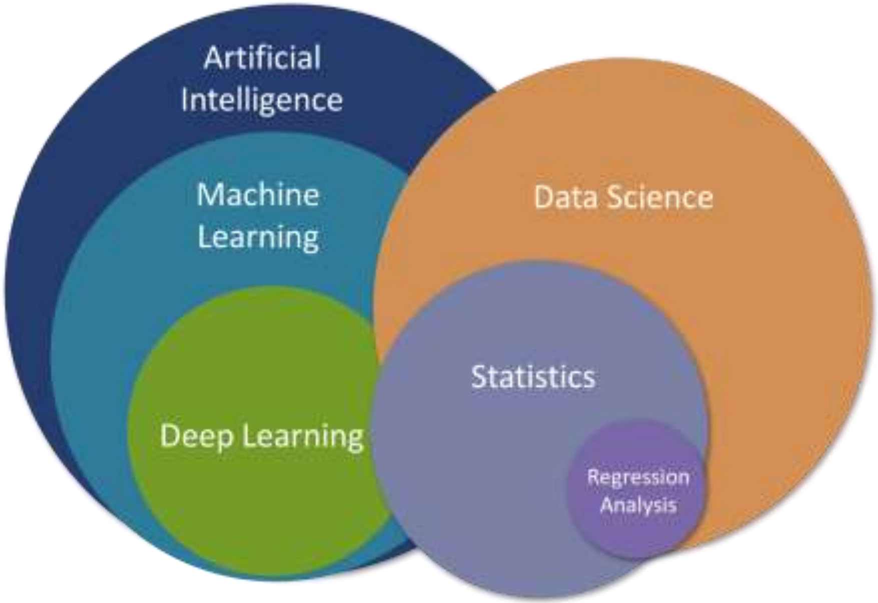
- Cambridge dictionary

Strong AI	**Weak** AI (Current Approach)
The machine must reason like a human (same mechanisms of operation).	The machine must arrive at the same solutions as a human (regardless of the method used).
Cognitive Approach	**Pragmatic Approach**

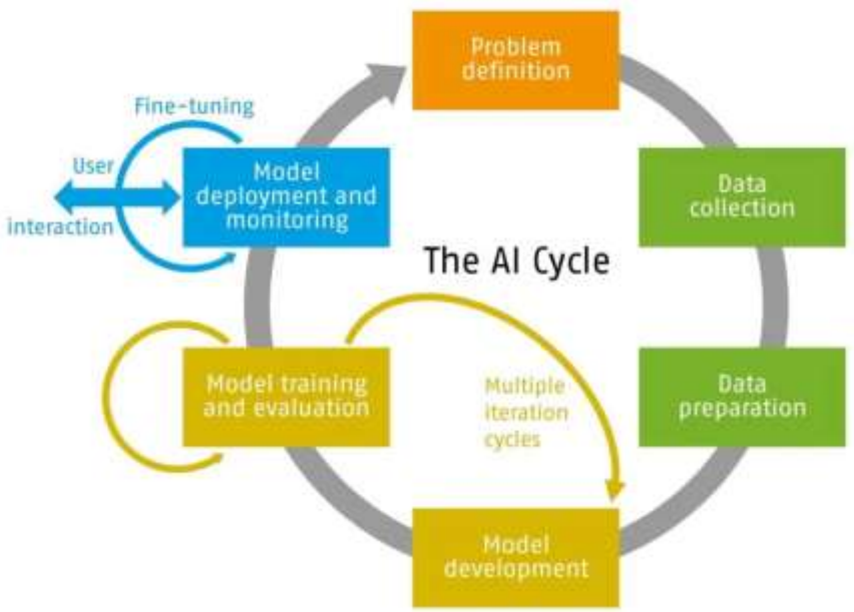
Source: [What is Artificial intelligence? - It's Working Explained]

Related Fields

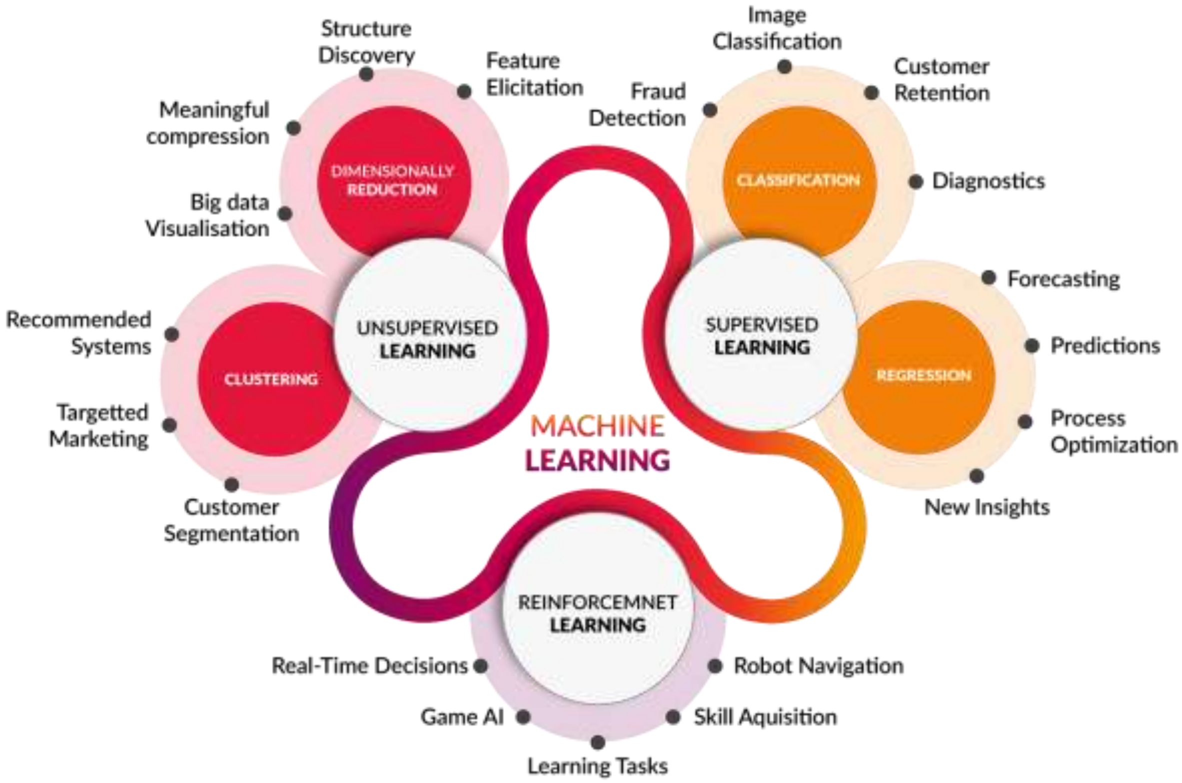




The AI (and Data Science) Cycle

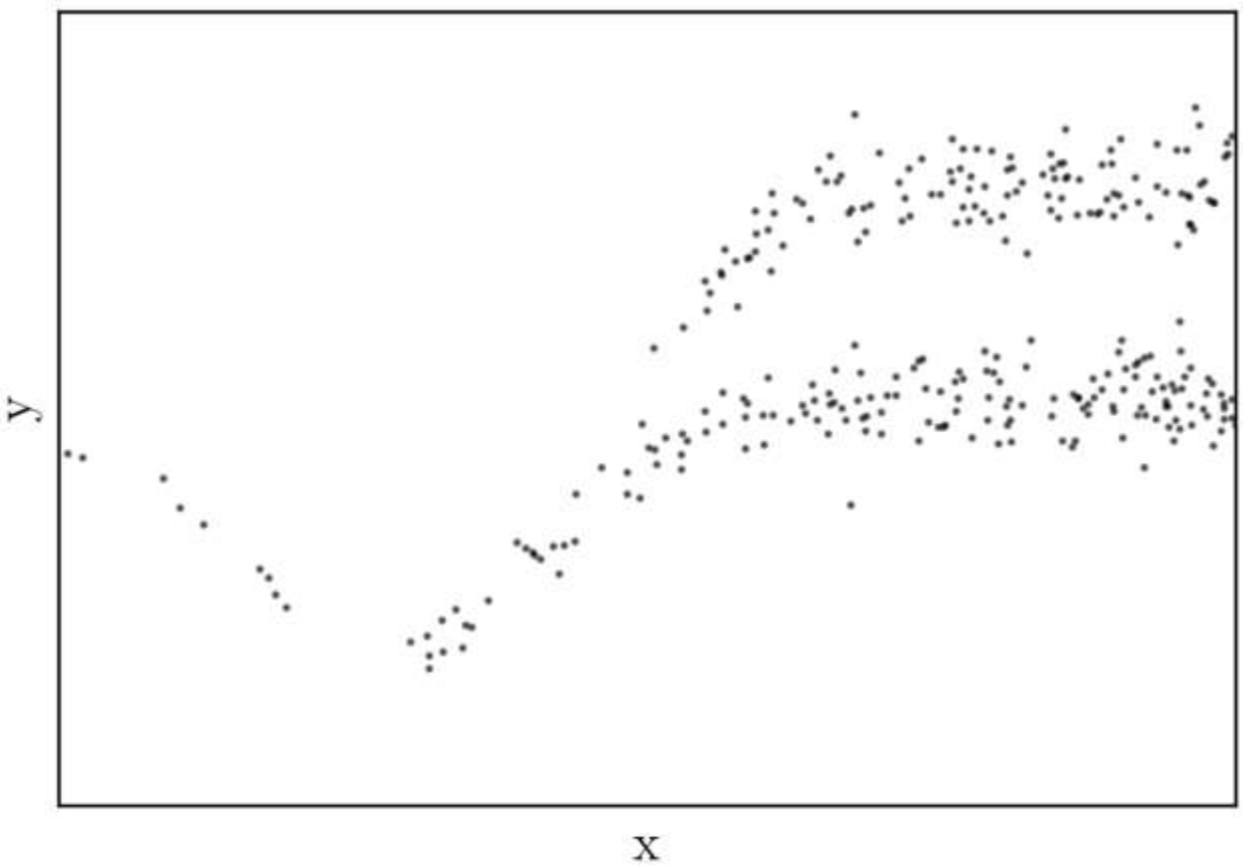


Learning Paradigms

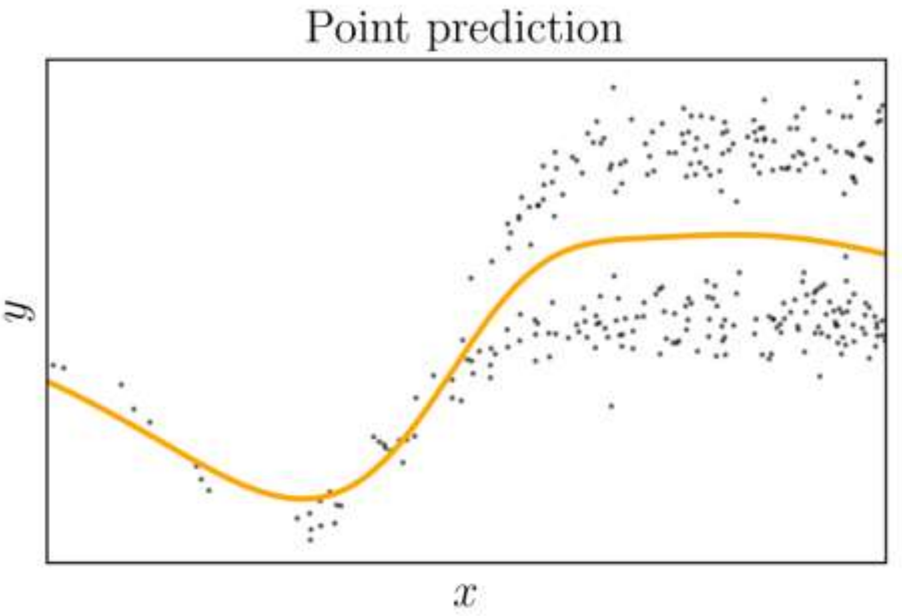


Source: [A Detailed and Understandable Explanation of Machine Learning]

Supervised learning - regression

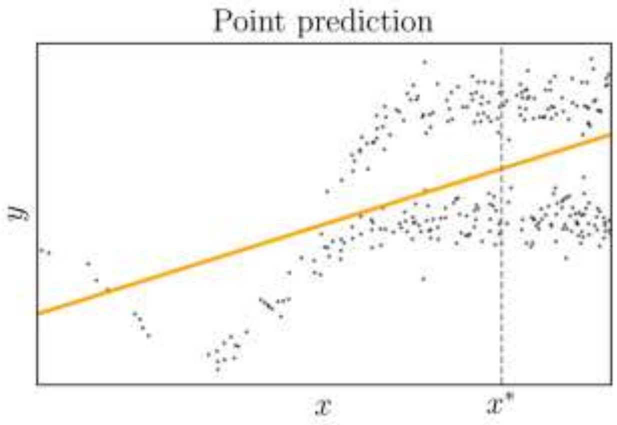


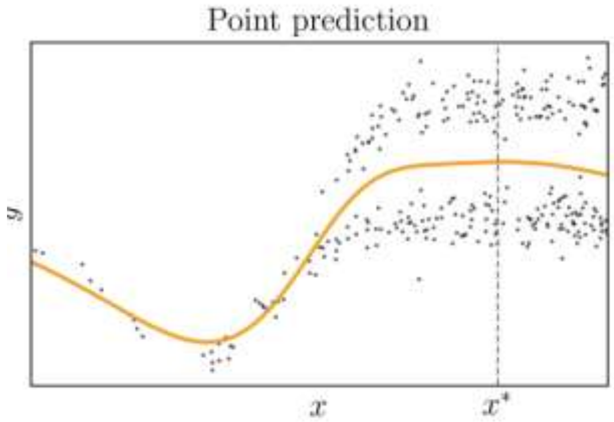
Deterministic (point) vs probabilistic predictions



- **Point** predictions: Single scalar, does not quantify uncertainty,

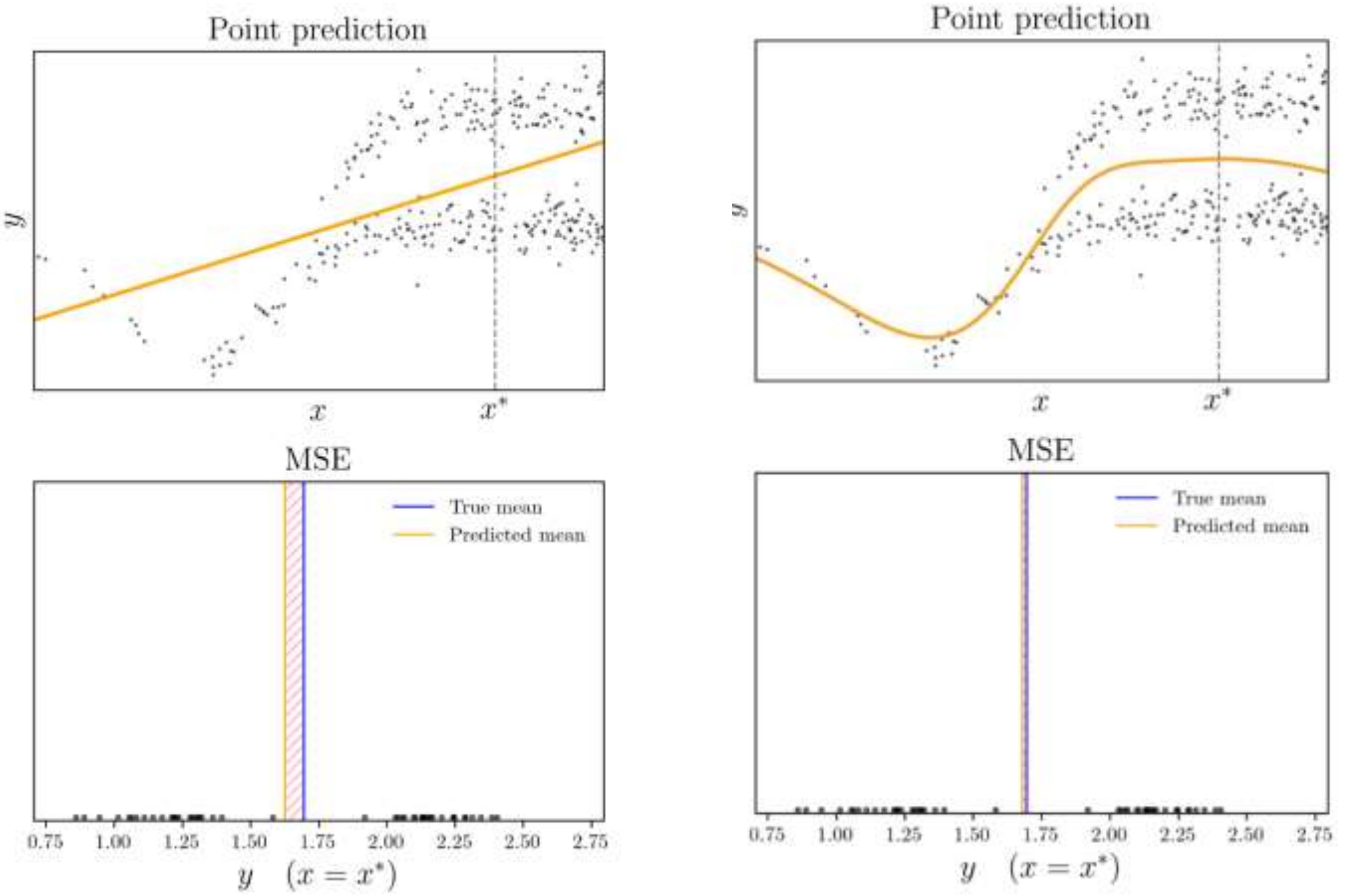
Deterministic (point) forecasts



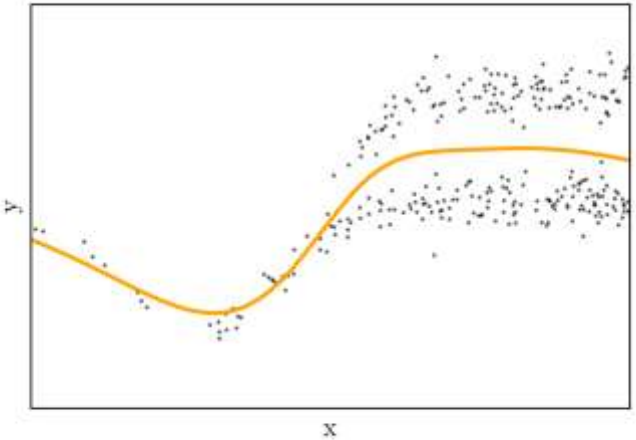


- **Deterministic** (point) predictions: Single scalar, does not quantify uncertainty,
- Which one is better, and why?

Deterministic (point) forecasts - MSE loss



MSE loss function



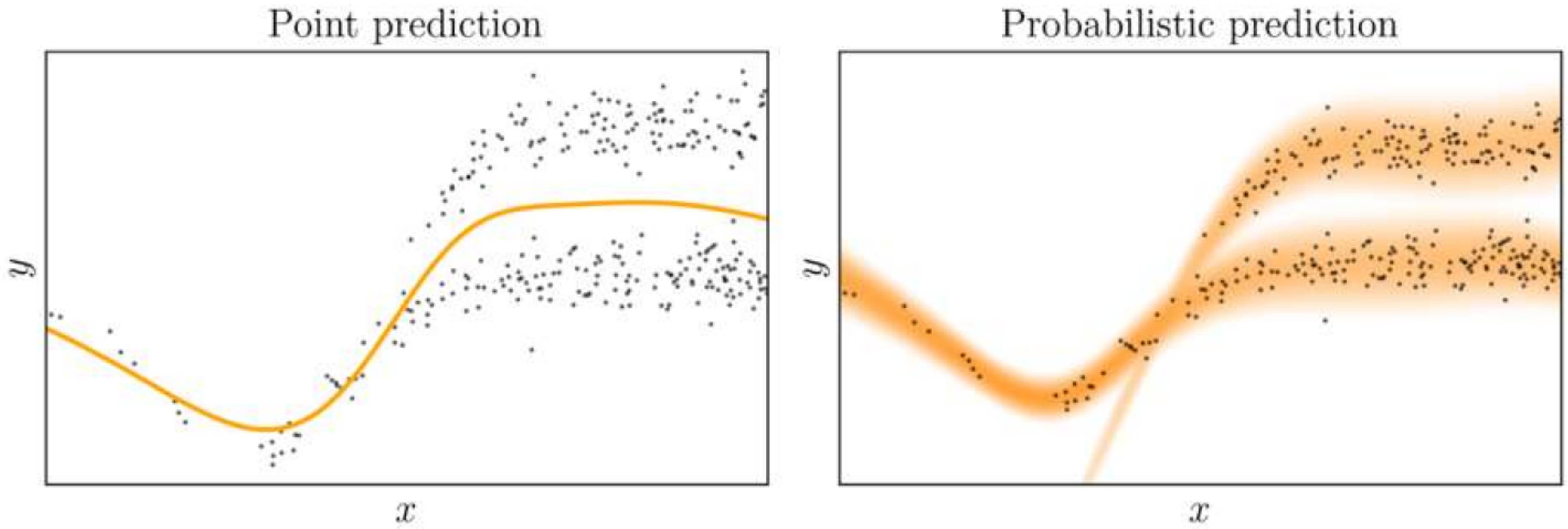
The mean squared error (MSE) compares a scalar-valued prediction $\hat{y}(x)$ with the true outcome Y given $X = x$:

$$\mathrm{MSE} = \mathbb{E}\left[(Y - \hat{y}(X))^2\right].$$

Minimizing the MSE leads to

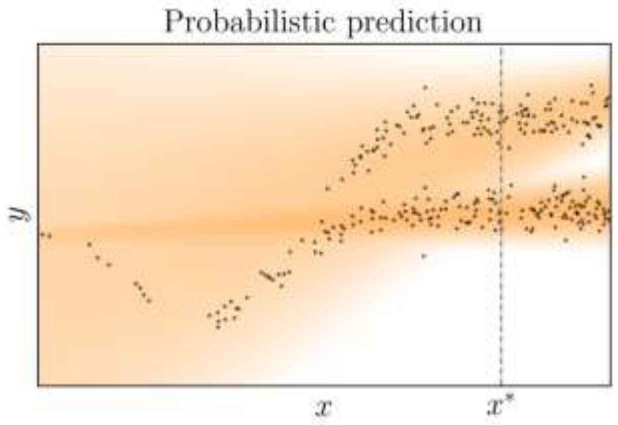
$$\hat{y}(x) \to \mathbb{E}[Y \mid X = x].$$

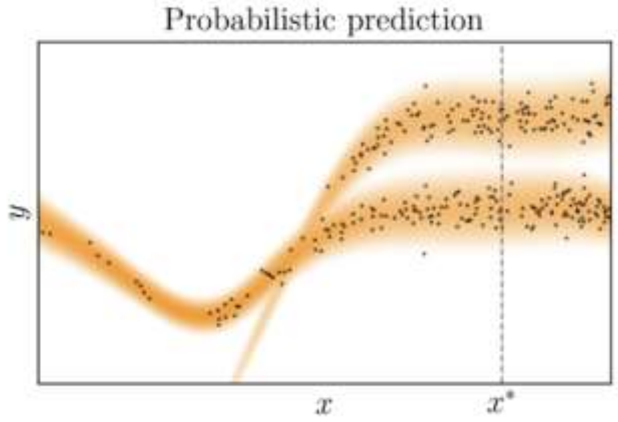
Deterministic (point) vs probabilistic forecasts



- Point predictions: Single scalar, does not quantify uncertainty,
- **Probabilistic** predictions: Probabilistic quantity, quantifies uncertainty.

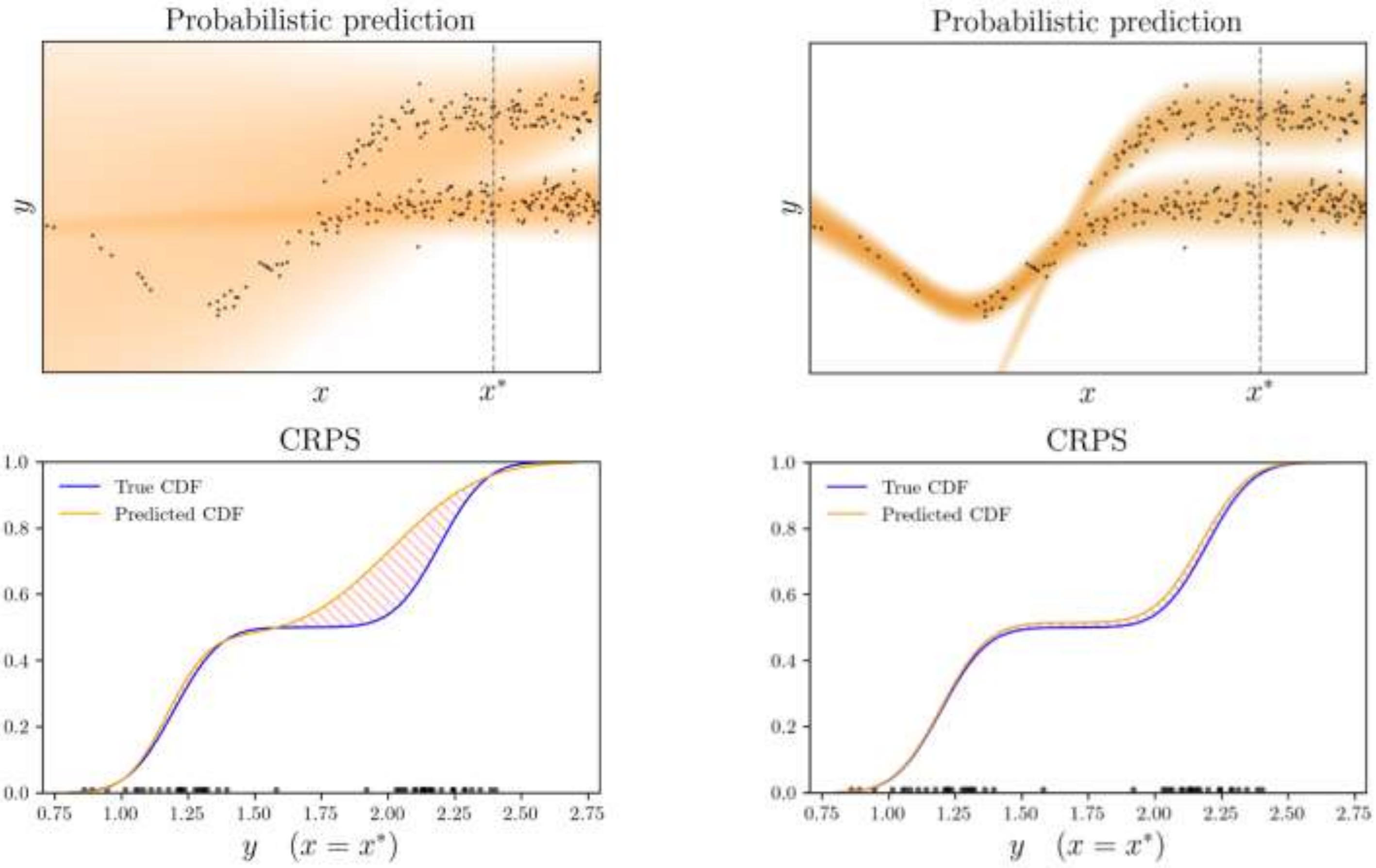
Probabilistic forecasts



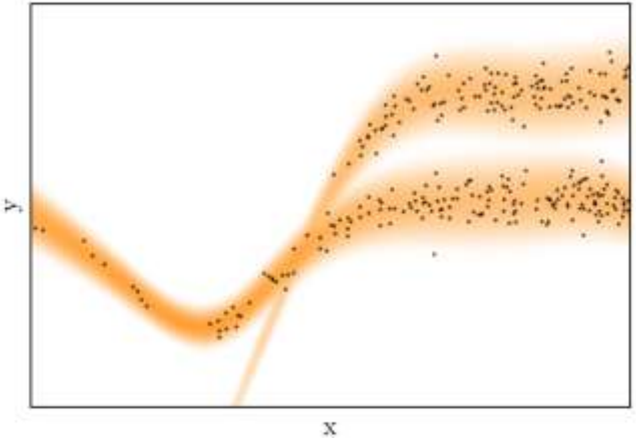


Which one is better, and why?

Probabilistic forecasts - CRPS loss



Loss functions: CRPS



The continuous ranked probability score (CRPS) compares a predicted CDF $\hat{F}(\cdot \mid x)$ with the true outcome Y given $X = x$:

$$\mathrm{CRPS}(\hat{F}(\cdot \mid x), Y) = \int_{-\infty}^{\infty} \left(\hat{F}(z \mid x) - \mathbf{1}\{Y \leq z\}\right)^2 dz.$$

Minimizing the CRPS leads to

$$\hat{F}(\cdot \mid x) \to \mathbb{P}(Y \leq \cdot \mid X = x).$$

Loss functions: CRPS vs fCRPS vs afCRPS

The CRPS for an ensemble forecast $\{z_j\}_{j=1}^M$ is given by

$$\mathrm{CRPS}\left(\{z_j\}_{j=1}^M, y\right) = \frac{1}{M}\sum_{j=1}^{M}|z_j - y| - \frac{1}{2M^2}\sum_{j=1}^{M}\sum_{k=1}^{M}|z_j - z_k| .$$

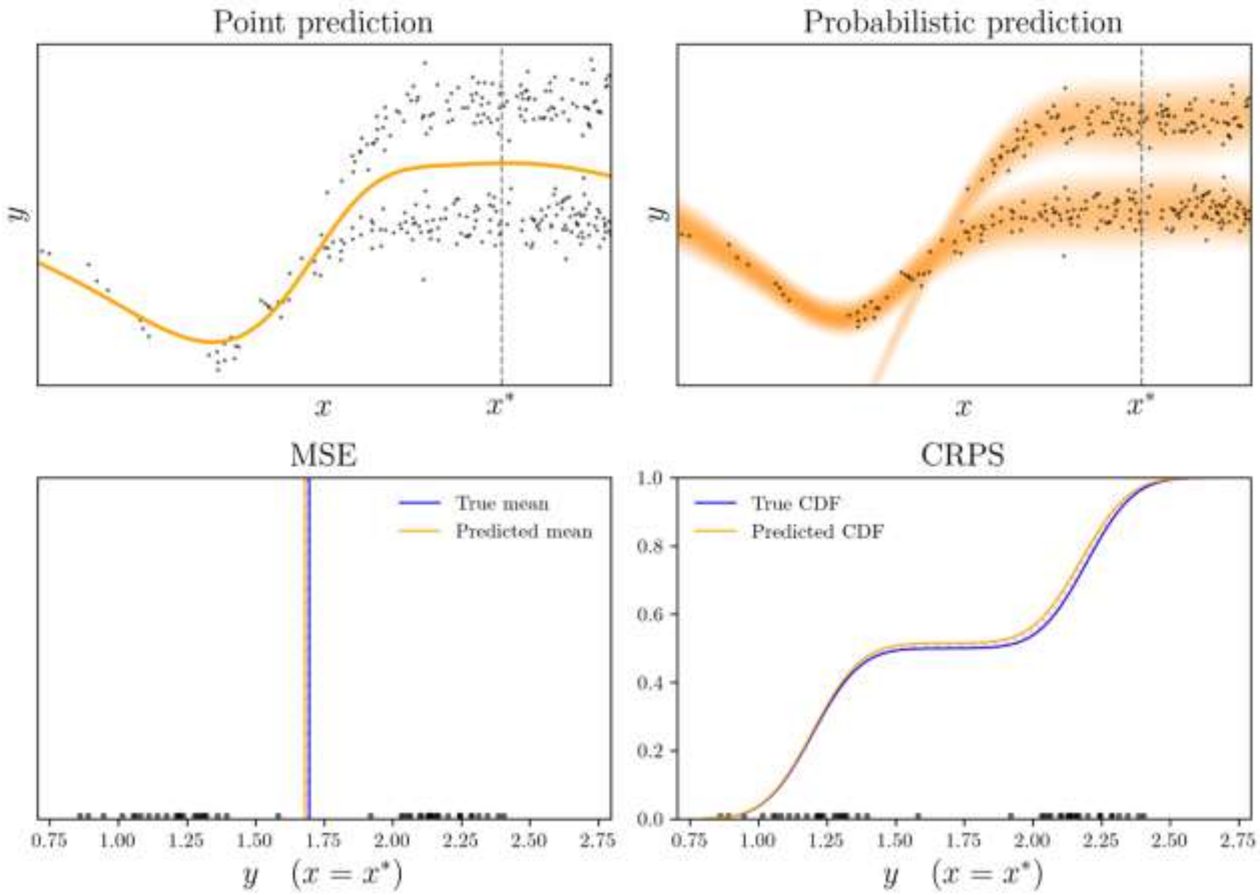
The *fair* CRPS (fCRPS) adjusts for ensemble size:

$$\mathrm{fCRPS}\left(\{z_j\}_{j=1}^M, y\right) = \frac{1}{M}\sum_{j=1}^{M}|z_j - y| - \frac{1}{2M(M-1)}\sum_{j=1}^{M}\sum_{k=1}^{M}|z_j - z_k| .$$

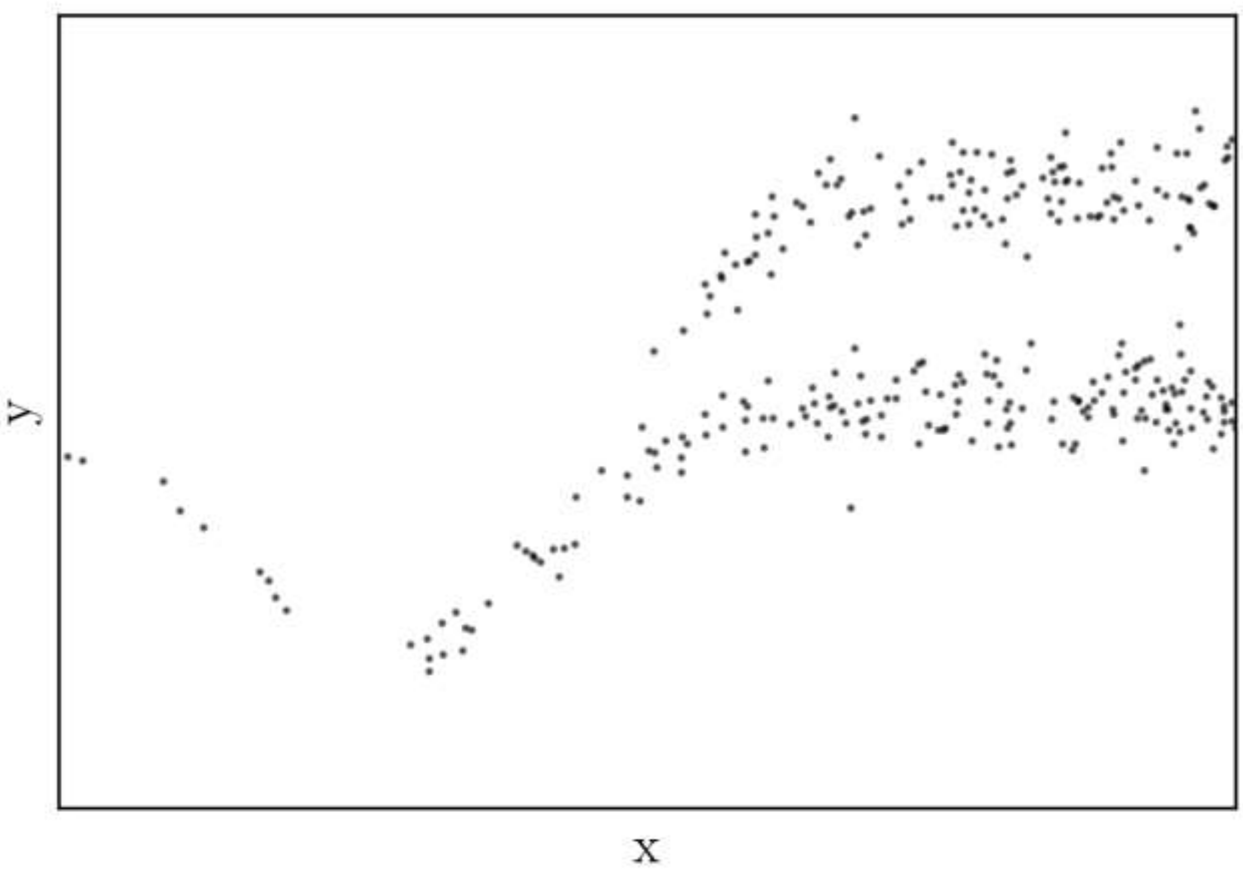
The *almost fair* CRPS (afCRPS) blends fCRPS and CRPS, fixing a degeneracy in case all members z_j apart from one are equal to y:

$$\mathrm{afCRPS}_\alpha = \alpha\,\mathrm{fCRPS} + (1-\alpha)\,\mathrm{CRPS}, \quad \alpha \in (0, 1].$$

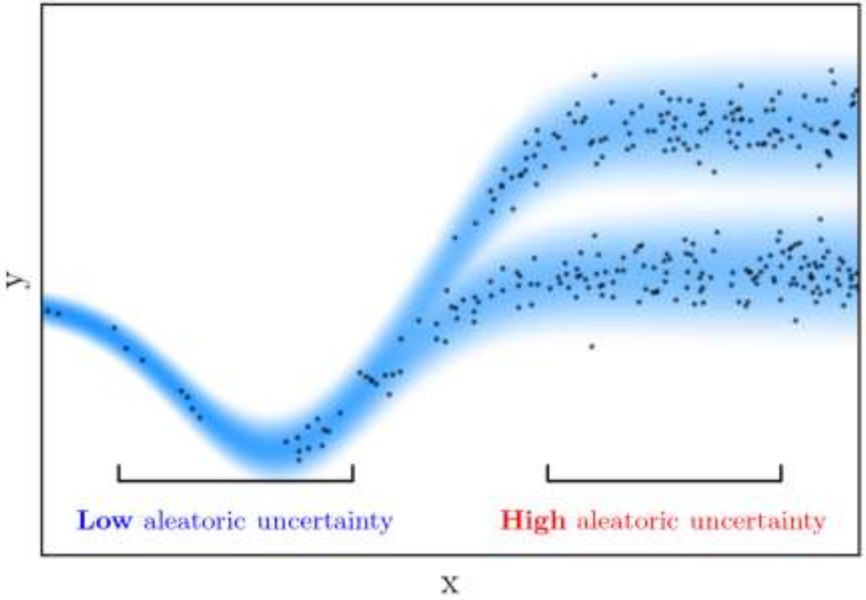
Loss functions: MSE vs CRPS



Sources of uncertainty?

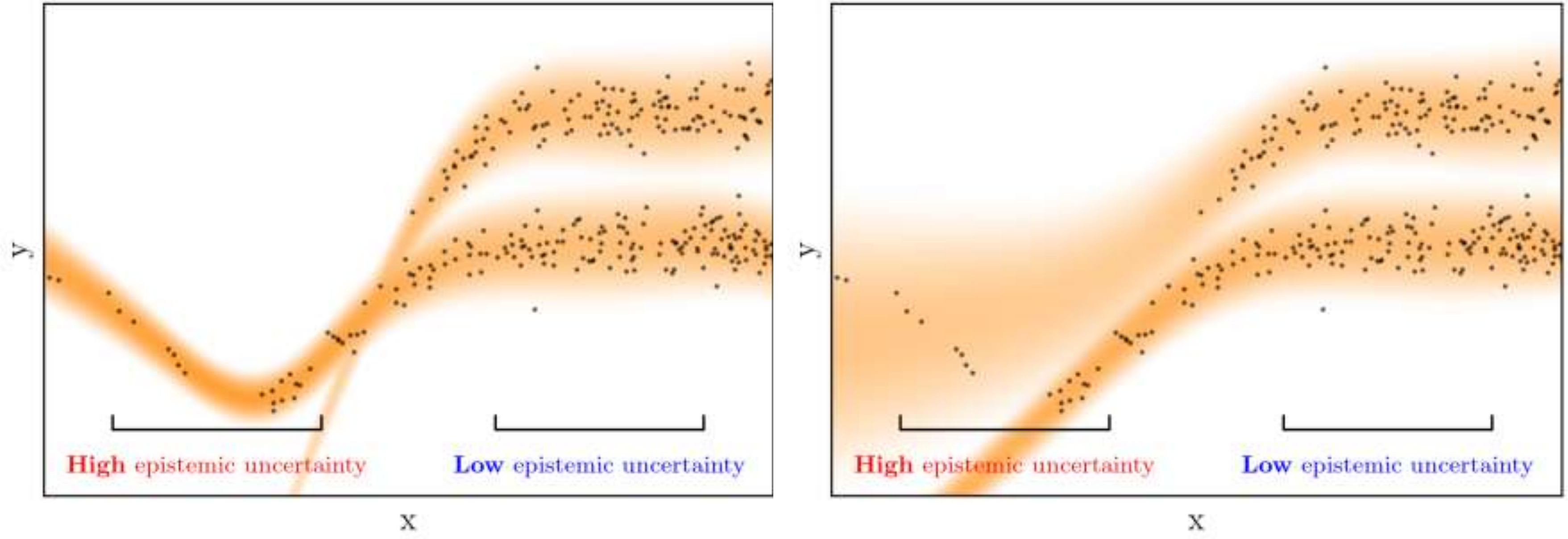


Sources of uncertainty



- **Aleatoric**: Irreducible, inherent to the data,

Sources of uncertainty

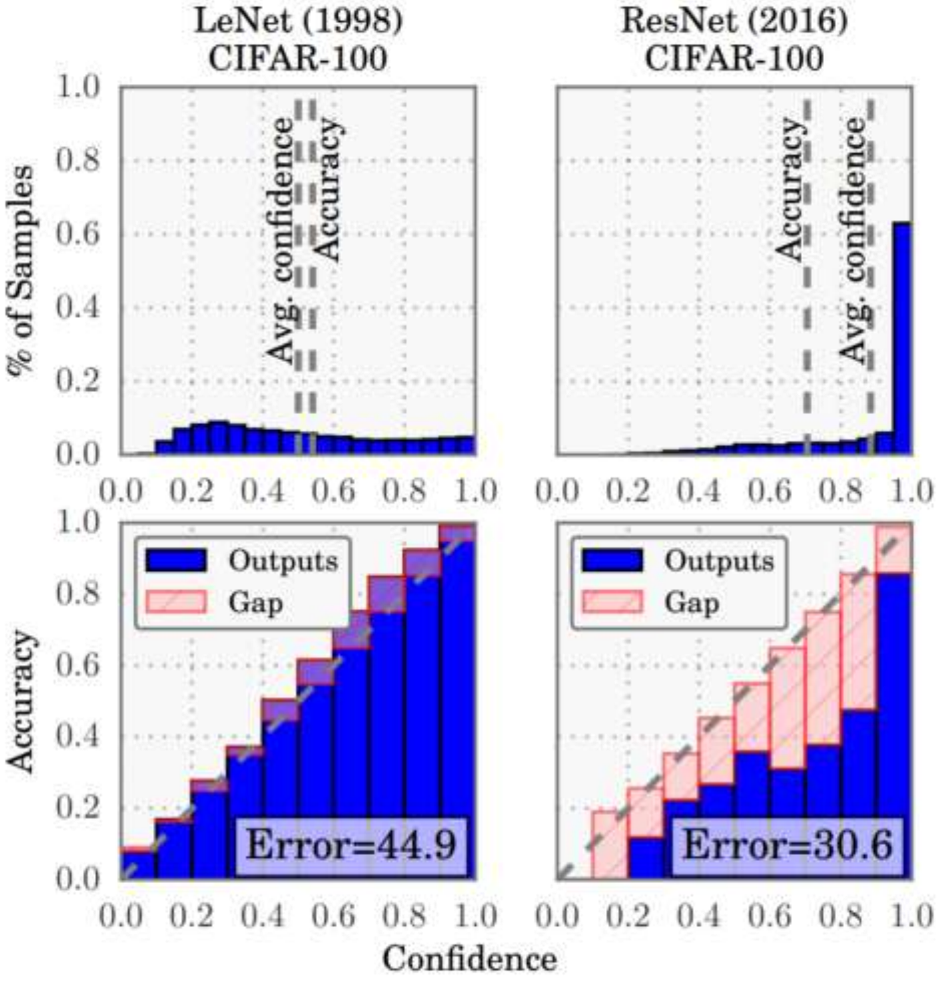


- **Aleatoric**: Irreducible, inherent to the data,
- **Epistemic**: All remaining uncertainty that arises in predictive modelling (lack of data, incorrect assumptions…).

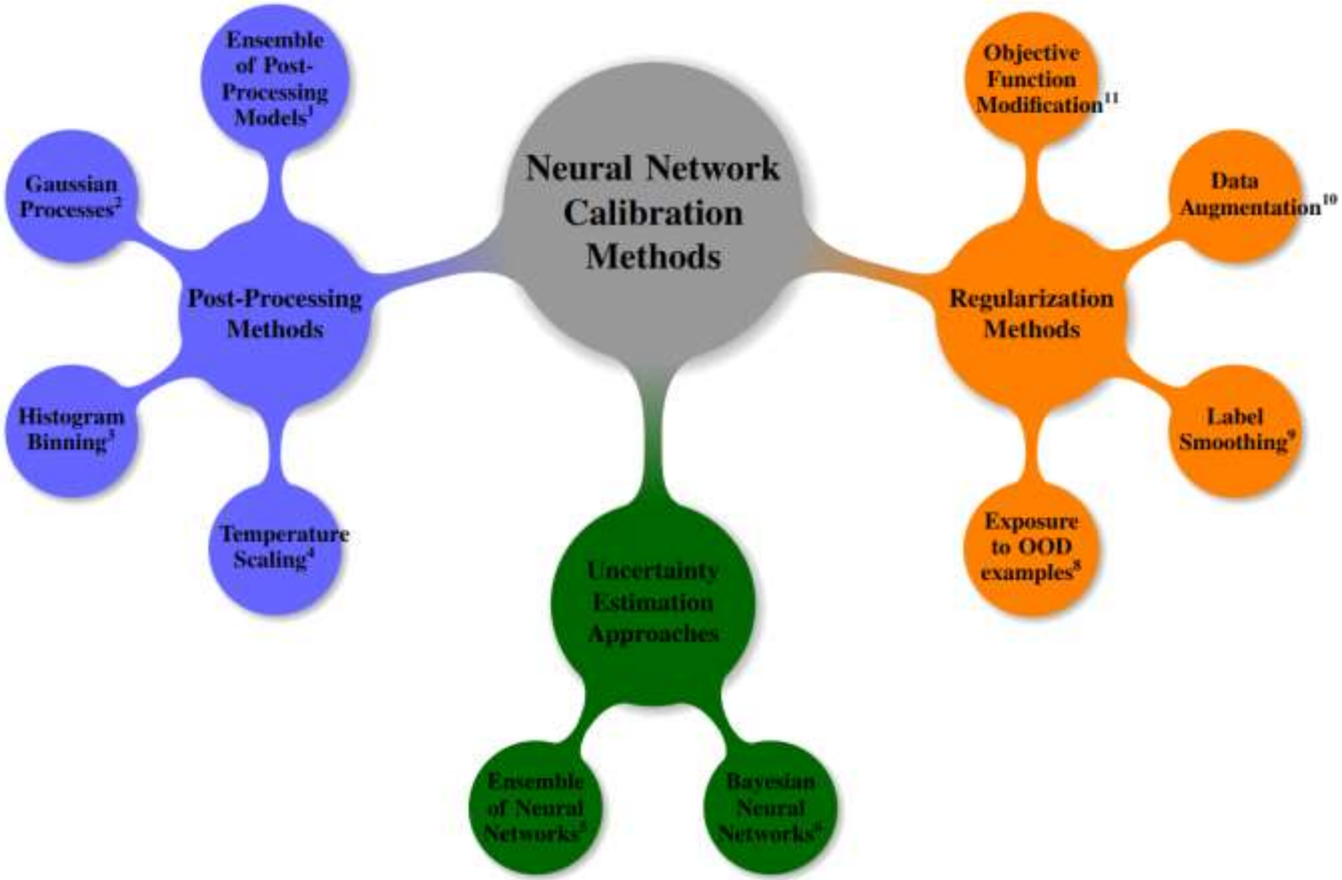
Probabilistic calibration

Neural network classifiers tend to be **overconfident**:

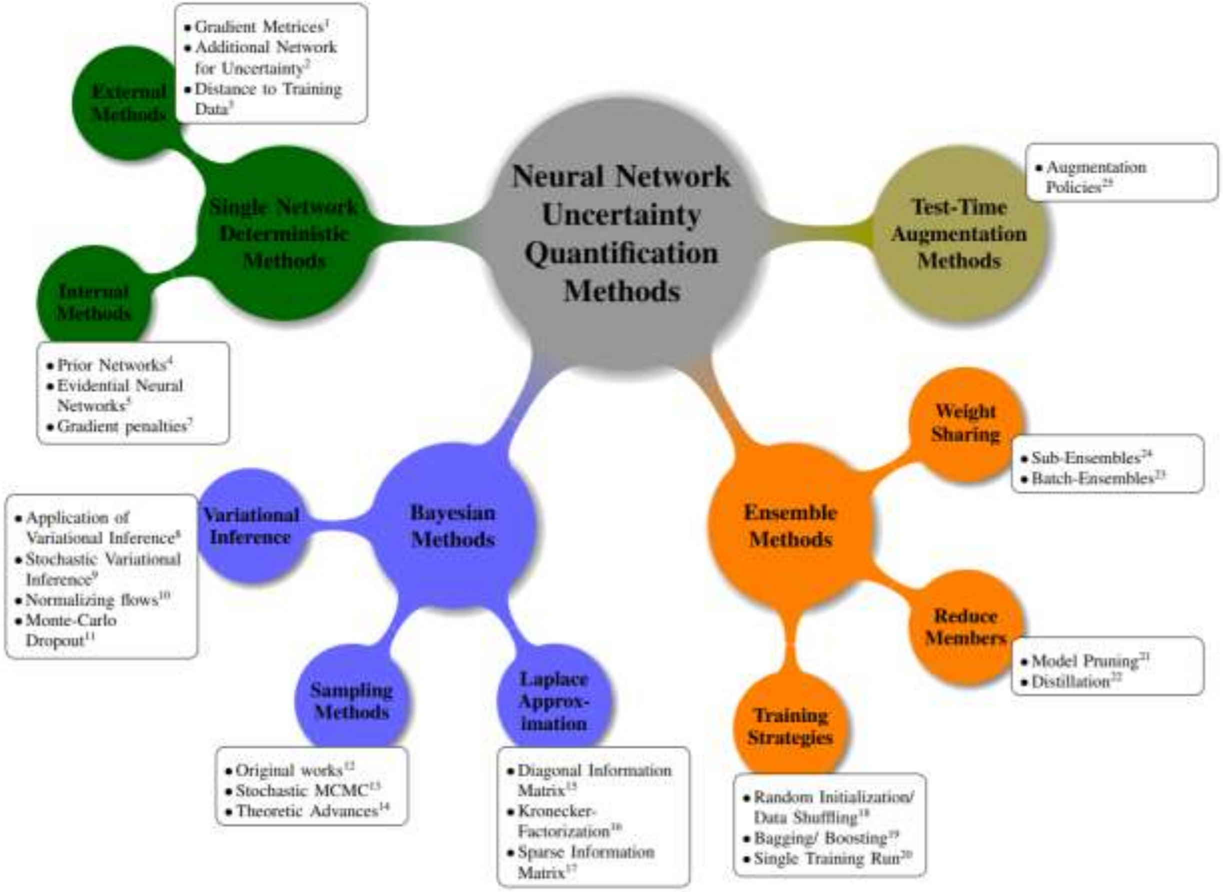
Predicted probabilities are larger than *observed frequencies.*



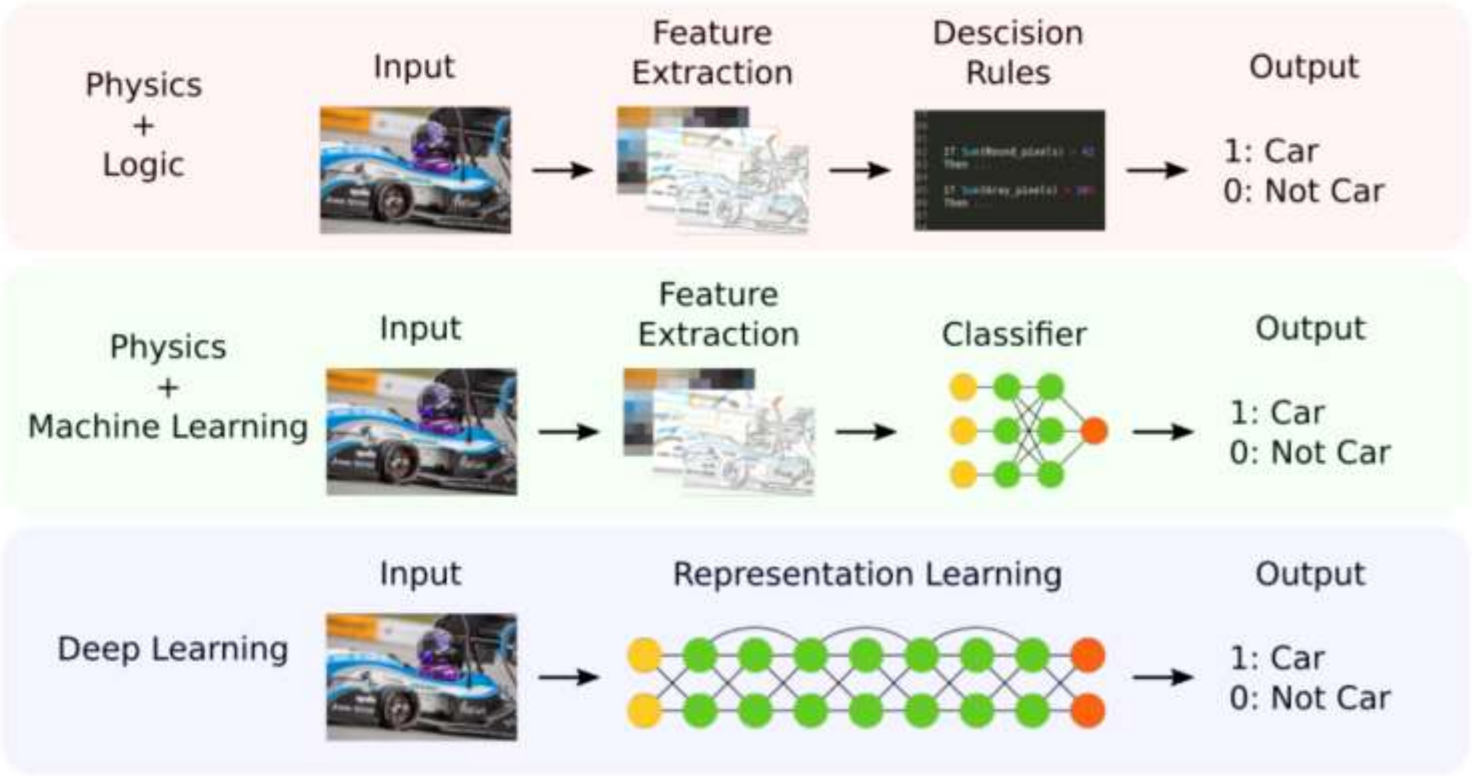
Neural Network Calibration Methods



Uncertainty Quantification Methods

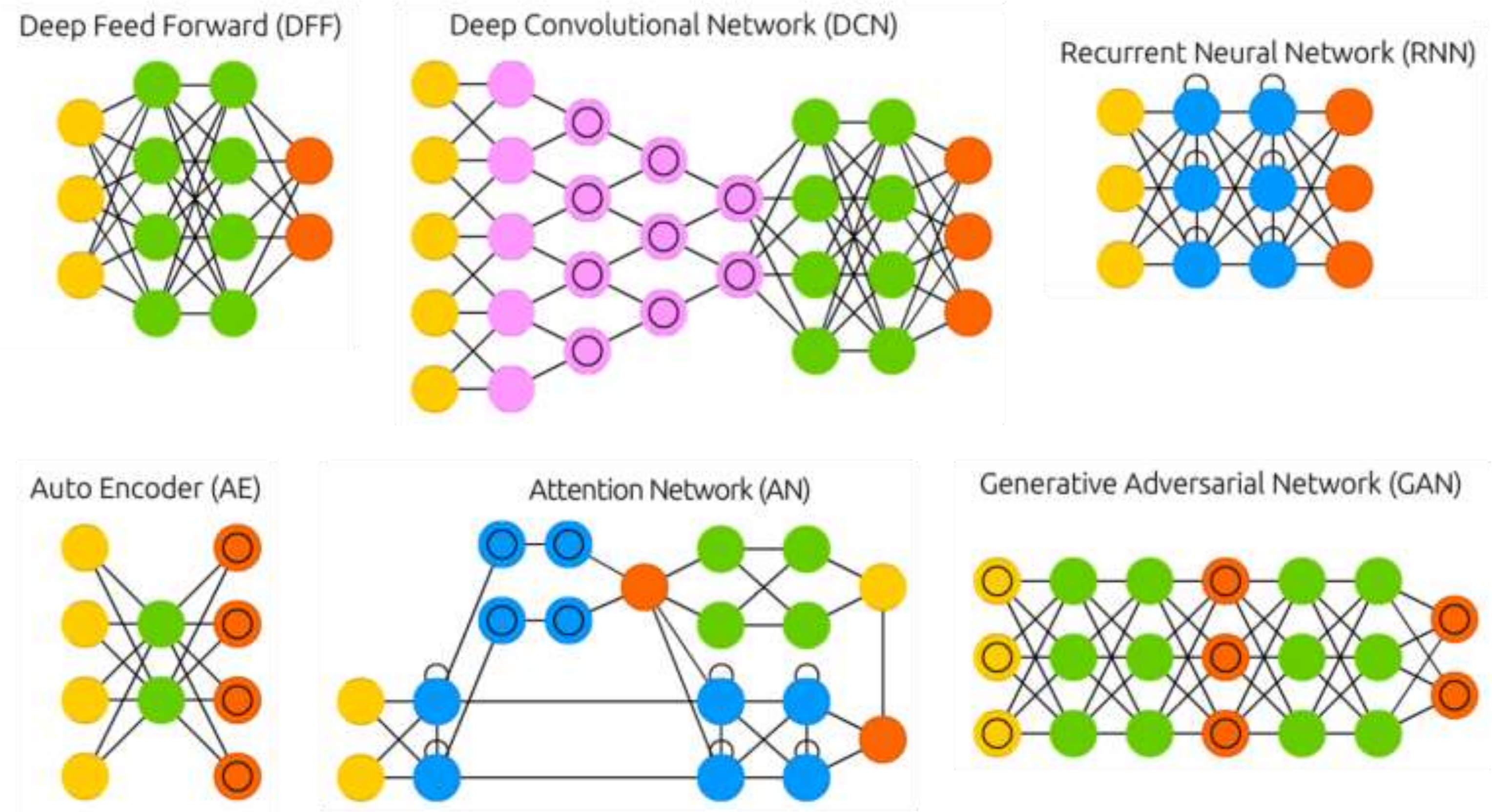


Deep Learning



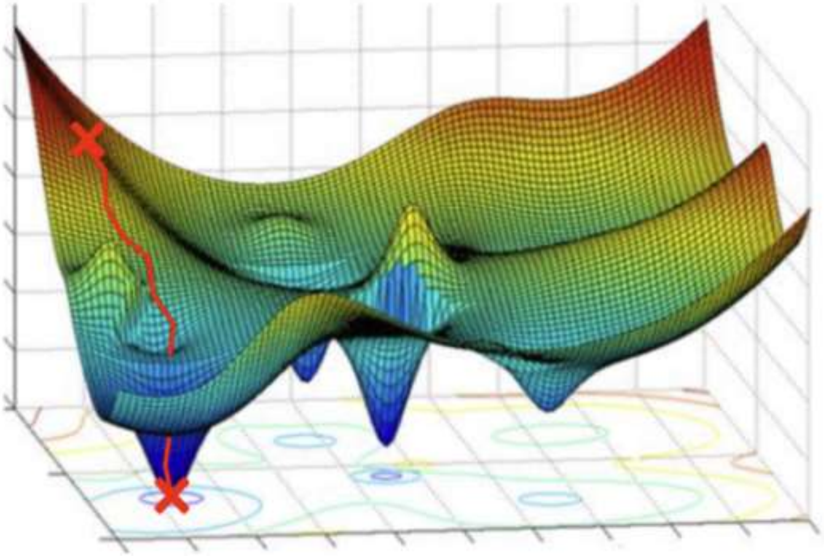
Source: [Deep Learning Course TRAIL]

Neural Network architectures

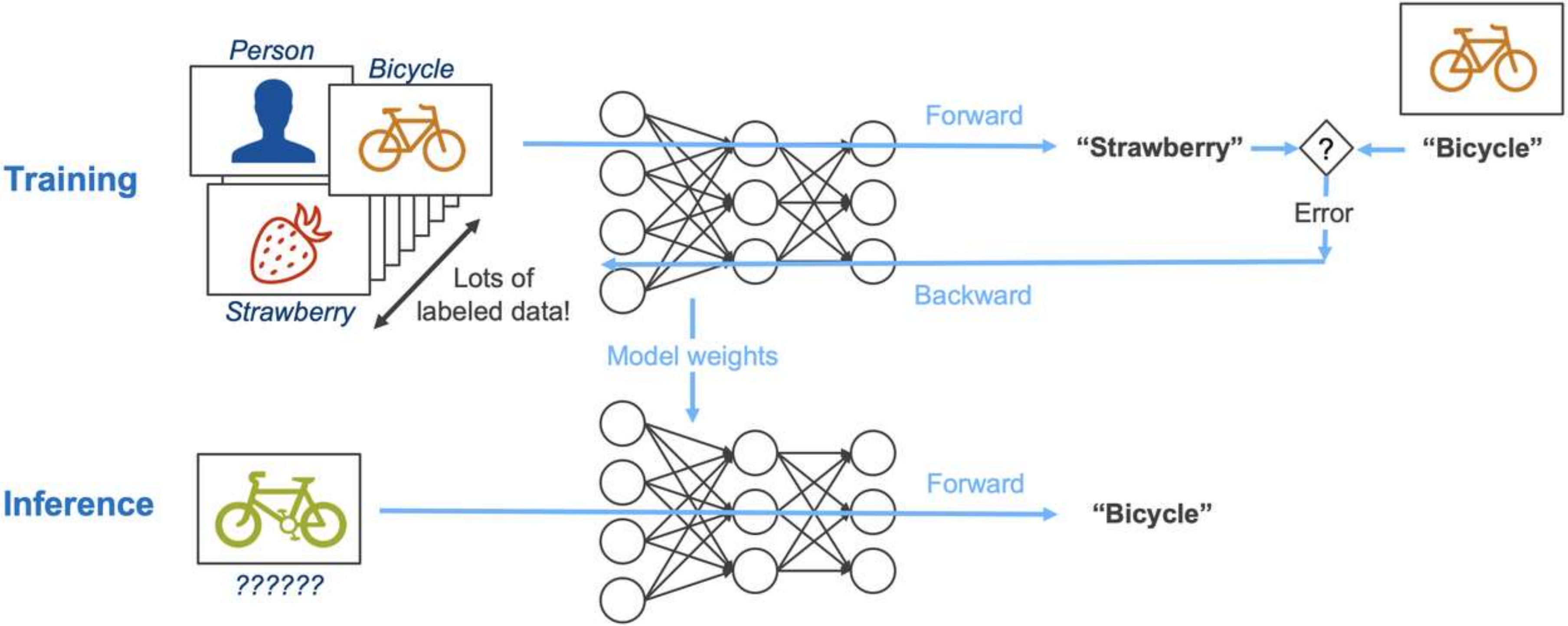


How do we train neural AI models (compute weights)?

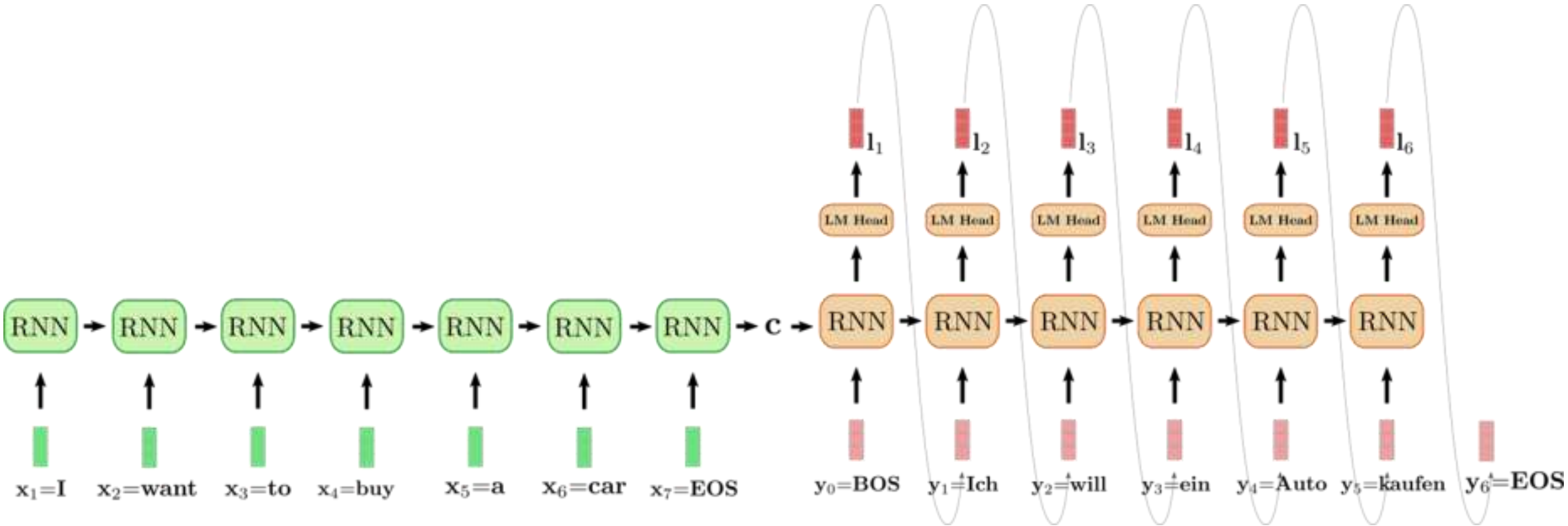
$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{CRPS}(F_{\theta}(\cdot \mid x_i), y_i).$$



Training vs inference

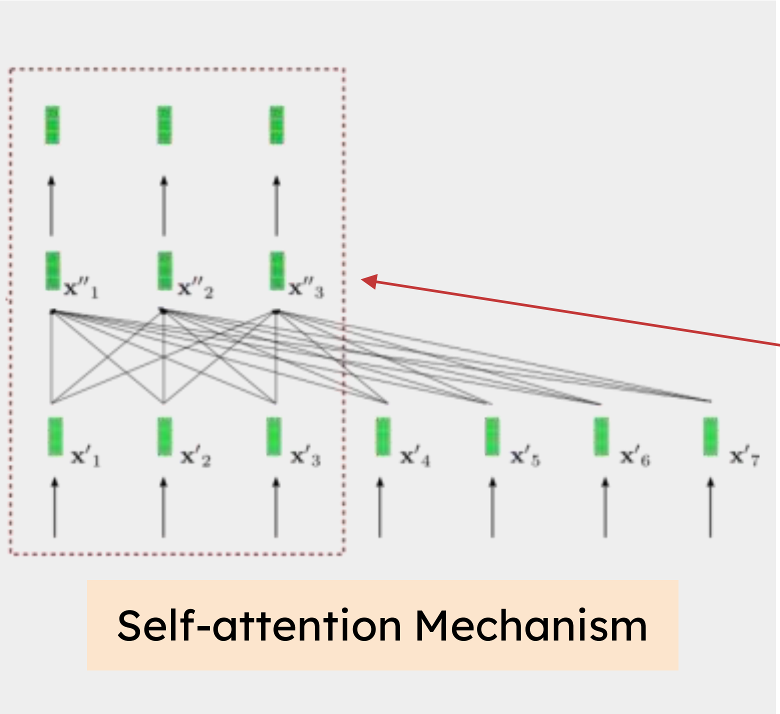


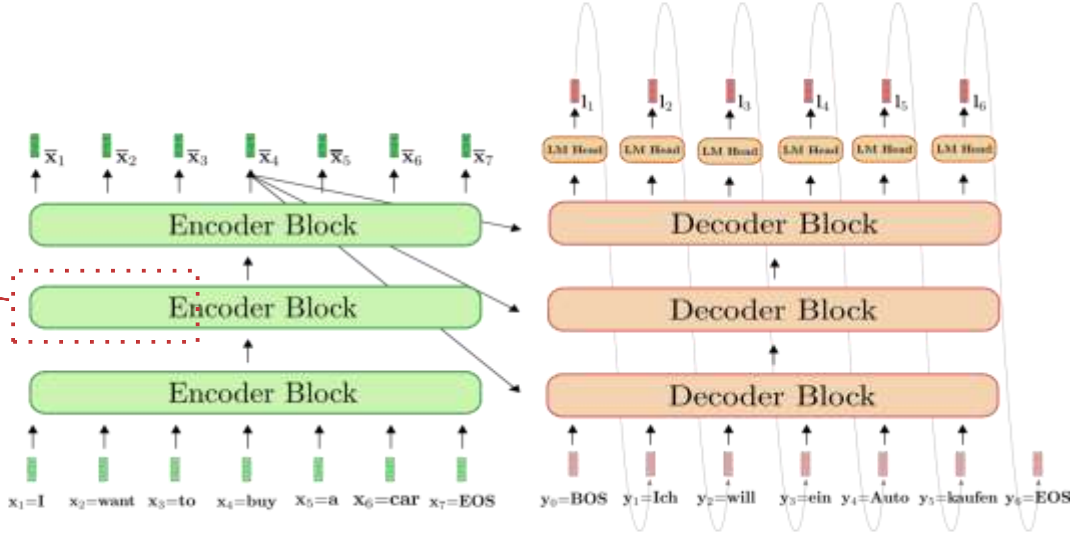
Autoregressive Encoder Decoder Models



Source: [Transformer-based Encoder-Decoder Models]()

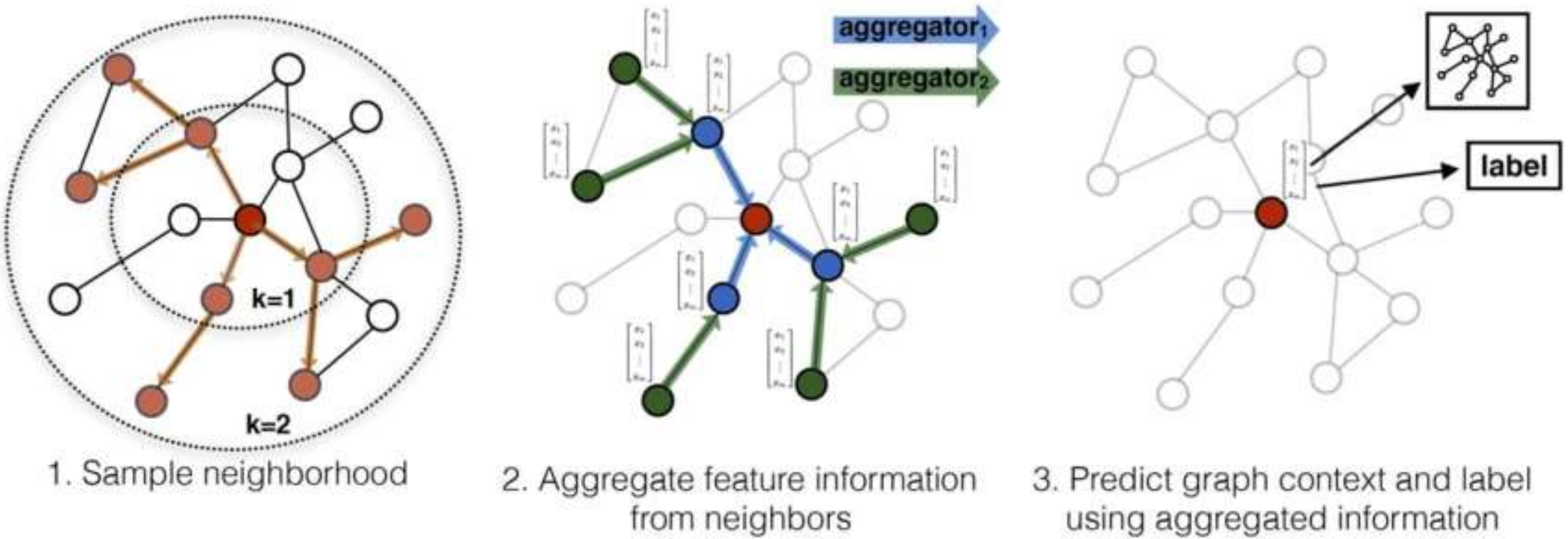
Transformers





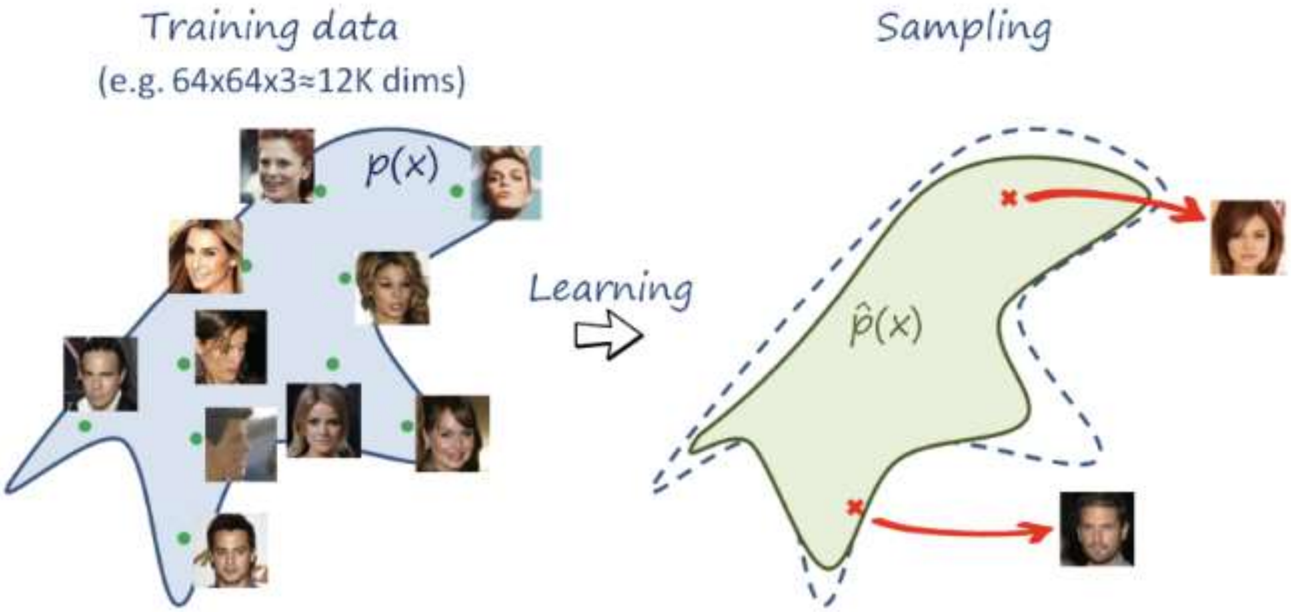
Source: [Transformer-based Encoder-Decoder Models]()

Graph Neural Networks

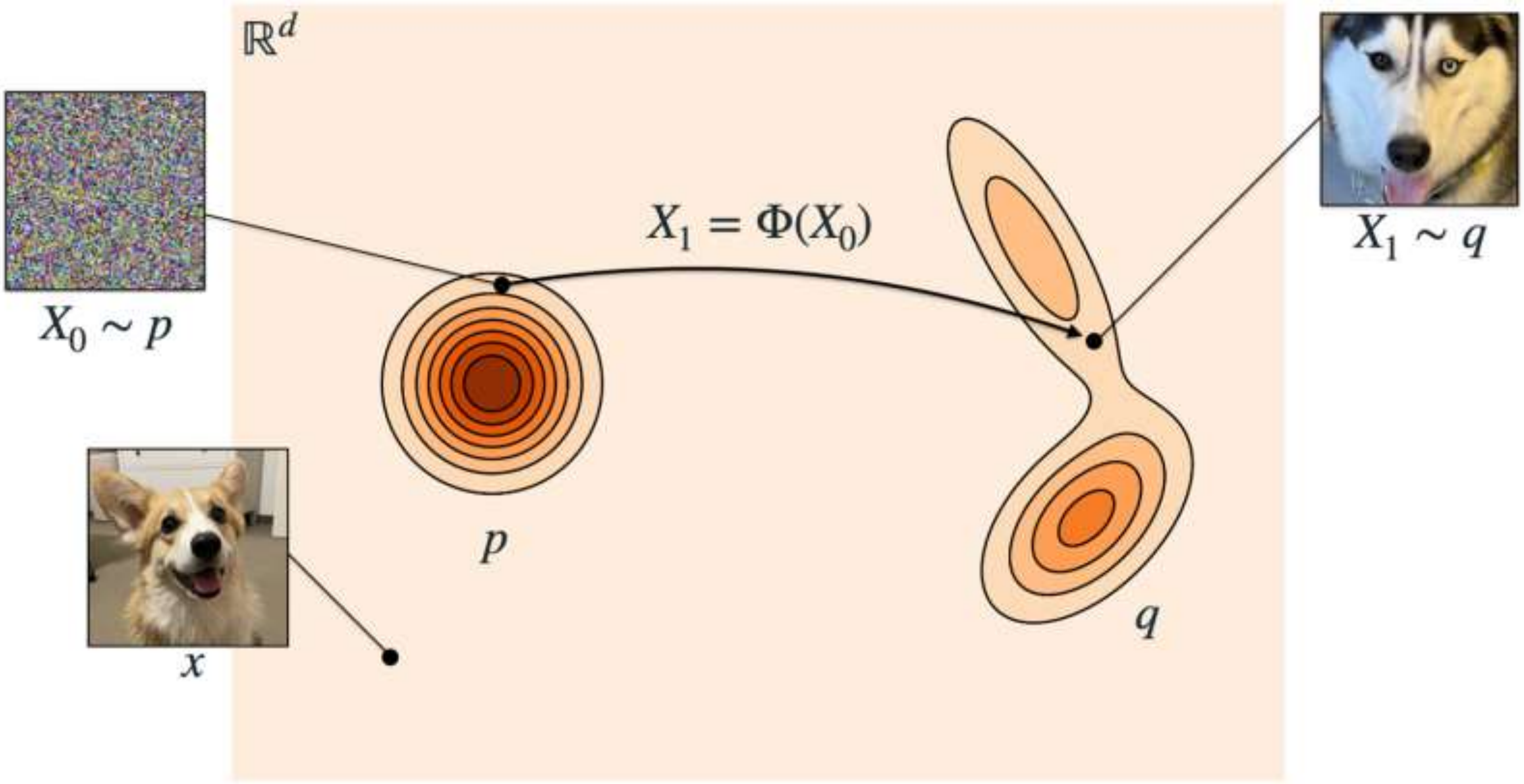


Source: [What Are Graph Neural Networks? | NVIDIA Blog]

Generative Models

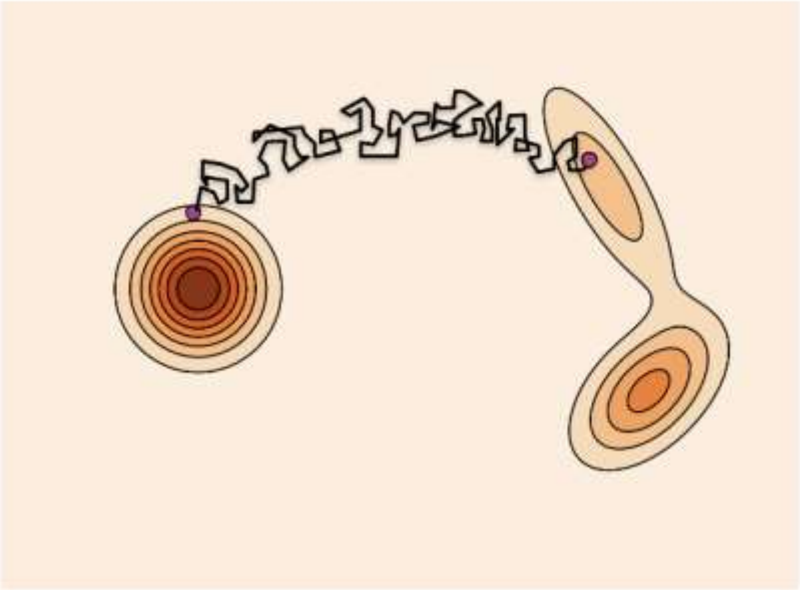


Generative Models



Source: [Flow Matching for Generative Modeling - NeurIPS Tutorial 2024]

Diffusion models

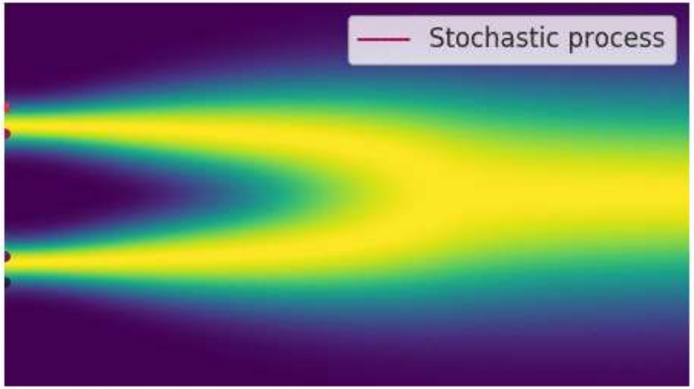


Source: [Flow Matching for Generative Modeling - NeurIPS Tutorial 2024]

Diffusion Models

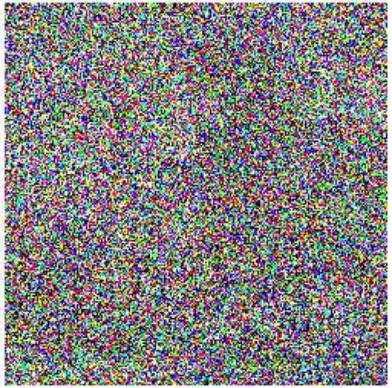
Forward Process (Fixed):

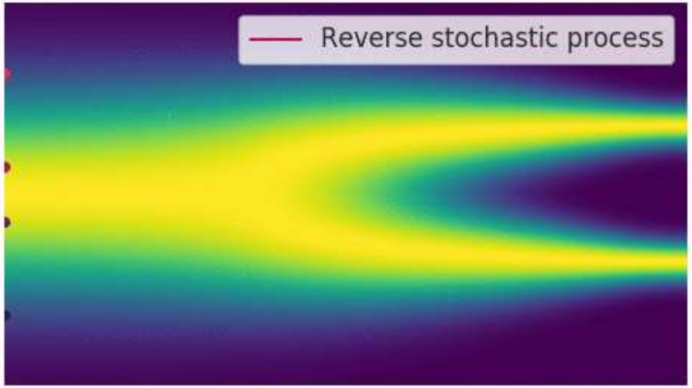
Destroy the data by gradually adding small amounts of Gaussian noise.



Reverse Process (Generative):

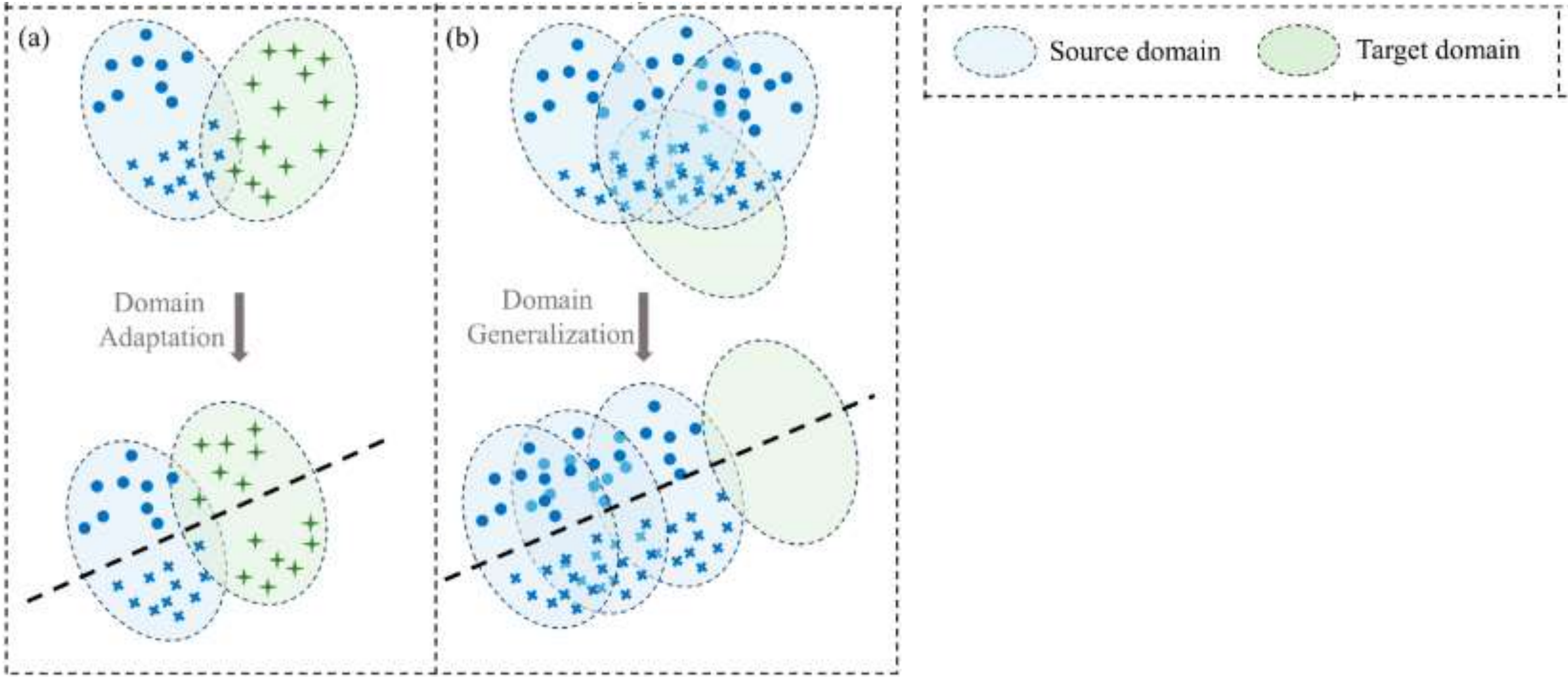
“Create” data by gradually denoising a noisy code from a stationary distribution.





Source: [Generative Modeling by Estimating Gradients of the Data Distribution | Yang Song]

Domain Adaptation and Generalization



Challenges

- **Data Quality & Availability:**
  - Sparse or biased data; missing measurements; historical records may be inconsistent
- **Generalization & Robustness**:
  - Models may overfit to past conditions; struggle with rare/extreme events (heatwaves, floods)
  - Complexity: Multi-scale spatial & temporal processes.
- **Uncertainty & Reliability:**
  - Difficult to quantify confidence in predictions; risk of overconfidence or underestimation
- **Physical consistency:**
  - Predictions may not adhere to the laws of physics.
- **Interpretability & Trust**:
  - “Black-box” predictions are hard to explain; decision-makers need transparency
- **Computational Demands:**