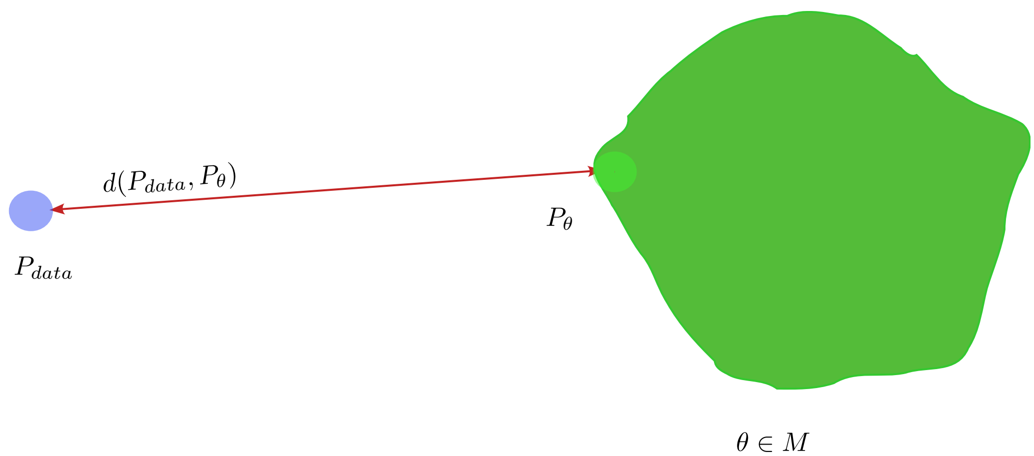
CS236 Lecture 2

May 7, 2024

1 The problem space



Model Family

- Over on the left, the blue represents the real world system from which we have data samples. It could be a set of images of dogs, for example. So $x_i \sim P_{data}$
- The green area is the space of possibilities of probability distributions that are parameterised by θ.
- We need to define the notion of distance or loss function d.

We want to learn a probability distribution $p(x)$ over images x such that:

- **Generation:** If we sample $x_{new} \sim p(x)$, x_{new} should look like a dog (*sampling*)
- **Density estimation:** $p(x)$ should be high if x looks like a dog, and low otherwise (*anomaly detection*)
- **Unsupervised representation learning:** We should be able to learn what these images have in common, e.g. ears, tails, etc. (*features*)

2 How do you represent the probability distribution $p(x)$?

For low dimensional data it would be straightforward. In the simplest case, binary distribution biased coin flip (Heads or Tails) is modelled by the Bernoulli Distribution.

- $D = \{Heads, Tails\}$
- Specify $P(X = Heads) = p$. Then $P(X = Tails) = 1 - p$
- Write $X \sim Ber(p)$
- Sampling: flip a (biased) coin

Extending this, we can have a Categorical distribution: a biased m-sided dice.

- $D = \{1, \cdots, m\}$
- $P(Y = i) = p_i$ such that $\sum p_i = 1$
- Write $Y \sim Cat(p_1, \cdots, p_m)$
- Sampling: roll a biased die