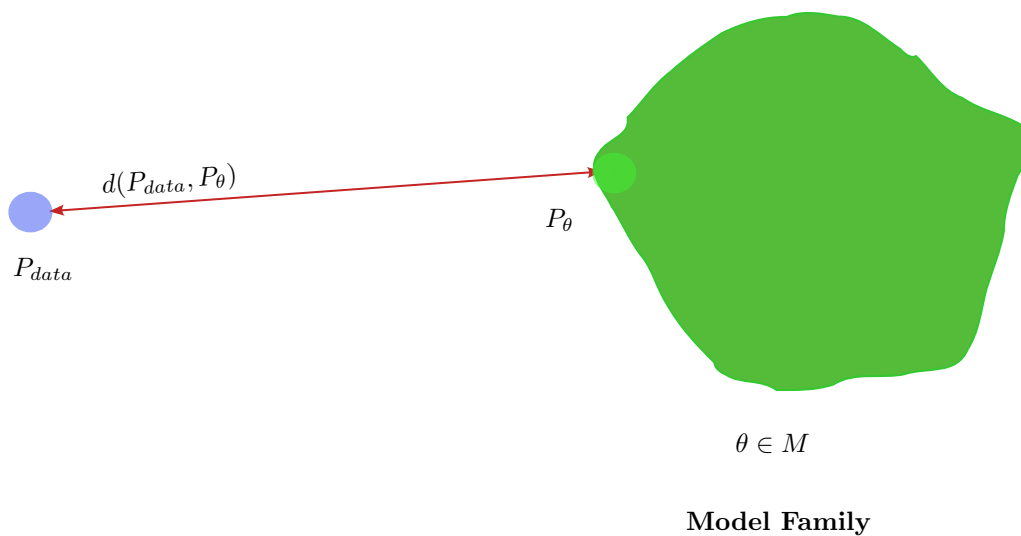


CS236 Lecture 2

May 7, 2024

1 The problem space



- 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.

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

2.1 Images

When we consider images, they are comprised of pixels. And each pixel can be considered three random variables, R, G, B each of which is taken from $\{0, \dots, 255\}$. Sampling from the joint distribution $(r, g, b) \sim p(R, G, B)$ randomly generates a colour for the pixel. The number of parameters to represent this space is $256 * 256 * 256 - 1$. We have the -1 because the probabilities we know have to add up to one, so the last parameter is implicitly known by subtracting from 1 the sum of all the other parameters. The problem is that for modest images, the number of parameters is too large. E.g. 28 x 28 pixels black and white for representing a handwritten digit means $2^{529} - 1$ parameters. We need to have some simplifying assumptions.

3 Structure through independence