Deep Learning

2.1 Quick visit to ML concepts

Dr. Konda Reddy Mopuri
kmopuri@iittp.ac.in
Dept. of CSE, IIT Tirupati
Aug-Dec 2021

Machine Learning

1. General objective is to capture the data regularity for making predictions

Machine Learning

1. General objective is to capture the data regularity for making predictions
2. In our linear regression example: we modeled the x and y as linearly related for fitting a line and predict y from x

Machine Learning

1. General objective is to capture the data regularity for making predictions
2. In our linear regression example: we modeled the x and y as linearly related for fitting a line and predict y from x
3. Broadly these are types of the inferences
   - Regression (e.g. customer satisfaction, stock prediction, etc.)
   - Classification (e.g. object recognition, speech processing, disease detection etc.)
   - Density estimation (e.g. sampling/synthesize, outlier detection, etc.)

Standard formalization

1. Classification and Regression: considers a measure of joint probability density $f_{X,Y}$ over the observation/value of interest and training (i.i.d.) samples $(x_n, y_n), n = 1, \ldots N$

Standard formalization

1. Classification and Regression: considers a measure of joint probability density $f_{X,Y}$ over the observation/value of interest and training (i.i.d.) samples $(x_n, y_n), n = 1, \ldots N$
2. Density estimation: distribution f_X and $x_n, n = 1, \ldots N$

Intuitive interpretation

1. Classification: $f_{X,Y}(x,y) = f_{X/Y=y}(x)P(Y=y)$

Intuitive interpretation

1. Classification: $f_{X,Y}(x,y) = f_{X/Y=y}(x)P(Y=y)$
2. Draw Y first, then given the value of Y generate X

Intuitive interpretation

1. Classification: $f_{X,Y}(x, y) = f_{X/Y=y}(x)P(Y = y)$
2. Draw Y first, then given the value of Y generate X
3. Conditional distribution $f_{X/Y}$ stands for the distribution of observable signal for category y (e.g. image of a dog, weight of a 30 year Indian male)

Intuitive interpretation

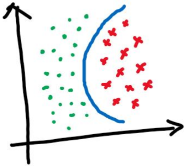
1. Regression: $f_{X,Y}(x,y) = f_{Y/X=x}(y) f_X(x)$

Intuitive interpretation

1. Regression: $f_{X,Y}(x,y) = f_{Y/X=x}(y)f_X(x)$
2. First generate X, given its value generate Y

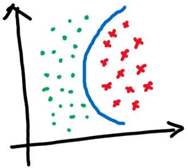
Summary: three types of inferences

1. Classification
   - X, Y random variables on $\mathcal{L} = \mathcal{R}^D \times \{1, \ldots, C\}$
   - Aim is to estimate the $\text{argmax}_y \; P(Y = y/X = x)$

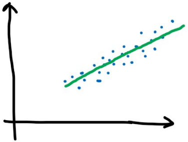


Summary: three types of inferences

1. Classification
   - X, Y random variables on $\mathcal{L} = \mathcal{R}^D \times \{1, \ldots, C\}$
   - Aim is to estimate the argmax$_y$ $P(Y = y/X = x)$

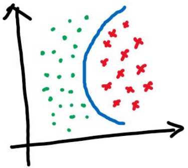


1. Regression
   - X, Y random variables on $\mathcal{L} = \mathcal{R}^D \times \mathcal{R}$
   - Aim is to estimate $\mathbb{E}(Y/X = x)$

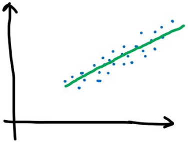


Summary: three types of inferences

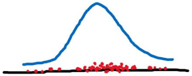
1. Classification
   - X, Y random variables on $\mathcal{L} = \mathcal{R}^D \times \{1, \ldots, C\}$
   - Aim is to estimate the $\text{argmax}_y \; P(Y = y/X = x)$



1. Regression
   - X, Y random variables on $\mathcal{L} = \mathcal{R}^D \times \mathcal{R}$
   - Aim is to estimate $\mathbb{E}(Y/X = x)$



1. Density estimation
   - X is random variable $\mathcal{R}^D$
   - Aim is to estimate the f_X



This categorization is not hard

Boundaries are vague

1. We may perform classification via class score regression

This categorization is not hard

Boundaries are vague

1. We may perform classification via class score regression
2. Density estimation can perform classification using Baye's rule

Risk/Loss

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$
2. How to find the goodness of a function f?

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$
2. How to find the goodness of a function f?
3. Through a loss $l : \mathcal{F} \times \mathcal{L} \rightarrow \mathcal{R}$

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$
2. How to find the goodness of a function f?
3. Through a loss $l : \mathcal{F} \times \mathcal{L} \rightarrow \mathcal{R}$
4. Such that value of $l(f, z)$ increases with the *wrongness* of f on z: (measure of discripency between the expected and predicted)

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$
2. How to find the goodness of a function f?
3. Through a loss $l : \mathcal{F} \times \mathcal{L} \rightarrow \mathcal{R}$
4. Such that value of $l(f, z)$ increases with the *wrongness* of f on z: (measure of discripency between the expected and predicted)
5.
   - Regression: $l(f, (x, y)) = (f(x) - y)^2$
   - Classification: $l(f, (x, y)) = \mathbf{1}(f(x) \neq y)$
   - Density estimation: $l(q, z) = -log(q(z))$

Loss/Risk

1. Learning: finding a good function f^* from a set of functions $\mathcal{F}$
2. How to find the goodness of a function f?
3. Through a loss $l : \mathcal{F} \times \mathcal{L} \rightarrow \mathcal{R}$
4. Such that value of $l(f, z)$ increases with the *wrongness* of f on z: (measure of discripency between the expected and predicted)
5. - Regression: $l(f, (x, y)) = (f(x) - y)^2$
   - Classification: $l(f, (x, y)) = \mathbf{1}(f(x) \neq y)$
   - Density estimation: $l(q, z) = -log(q(z))$
6. Loss may have additional terms (from prior knowledge)

Expected Risk

1. We want f with small *expected (average) risk* $R(f) = \mathbb{E}_z(l(f, z))$

Expected Risk

1. We want f with small *expected (average) risk* $R(f) = \mathbb{E}_z(l(f, z))$
2. $f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} R(f)$

Expected Risk

1. We want f with small *expected (average) risk* $R(f) = \mathbb{E}_z(l(f,z))$
2. $f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} R(f)$
3. This is unknown. However, if the training data $\mathcal{D} = \{z_1, \ldots, z_N\}$ is i.i.d. we can estimate the risk empirically (known as empirical risk),

   $\hat{R}(f; \mathcal{D}) = \hat{\mathbb{E}}_{\mathcal{D}}(l(f,z)) = \frac{1}{N} \sum_{i=1}^{N} l(f, z_n)$