CS6462

Probabilistic and Explainable AI

Lesson 15

Bayesian Nonparametric Models

*

Gaussian Process Regression Model

by Emil Vassev

February 24, 2024

$$p(\theta \mid y)$$

Bayesian Nonparametric Models

Definition:

- constitutes a Bayesian model on an infinite-dimensional parameter space
- the size of the model grows with data size
- parameter space - chosen as a set of all possible solutions to a given learning problem

Popular Models:

- Gaussian Process Regression - the correlation structure is refined with growing sample size
- Dirichlet Process Mixture Model for Clustering - adapts the number of clusters to the complexity of the data

Applications: variety of machine learning problems:

- regression, classification, clustering, latent variable modeling, sequential modeling, image segmentation, source separation, and grammar induction

$$p(\theta \mid y)$$

Bayesian Nonparametric Models (cont.)

$$P(\Theta|data) = \frac{P(data|\Theta) * P(\Theta)}{P(data)}$$

normalizing factor

Bayesian formulation of nonparametric problems:

- Bayes' Theorem – fixed set of parameters and observations
- nonparametric problems - dimension of the parameter space ***Θ*** should change with the sample size:
  - use an infinite-dimensional parameter space
  - invoke only a finite subset of the available parameters on any given finite data set
  - <u>infinite dimensional parameter space</u> – finite, but unbounded dimensional parameter space

Bayesian nonparametric model:

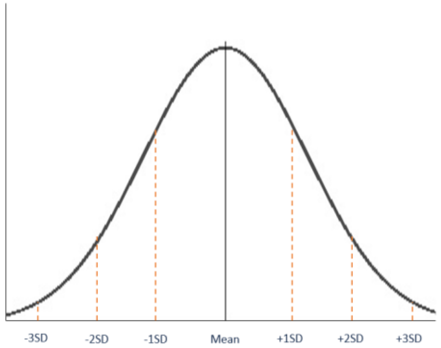
1) constitutes a Bayesian model on an infinite-dimensional parameter space
2) can be evaluated on a finite sample in a manner that uses only a finite subset of the available parameters to explain the sample

$p(\theta \mid y)$

Gaussian Process

Gaussian (Normal) distribution:

$P_X(x) = f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}$ where $x = \{-\infty, \ldots, +\infty\}$, $\mu = \{-\infty, \ldots, +\infty$

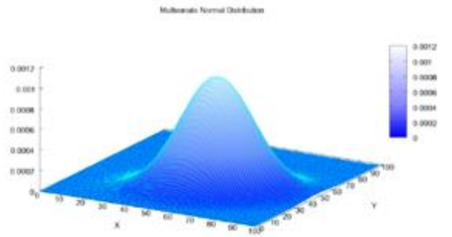


- X is random variable, μ is the mean (or expectation) of the distribution, and σ is the standard deviation of the distribution

Multivariate Normal (MVN) distribution:

- system is described by more random variables $X = \{X_1, X_2, \ldots, X_D\}$ that are <u>correlated</u>

$N(X;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D * |\Sigma|}} e^{\left\{-\frac{1}{2}*(X-\mu)^T * \Sigma^{-1} * (X-\mu)\right\}}$



- D is the dimension
- $\mu = \mathbb{E}[X] \in \mathbb{R}^D$ is the mean vector ($\mathbb{R}^D$ is the real coordinate space of dimension D)
- $\Sigma = cov[X]$ is a $D \times D$ covariance matrix
- $X = \{X_1, X_2, \ldots, X_D\}^T$ is multivariate random variable (random vector), or its transpose vector

$p(\theta \mid y)$

Gaussian Process (cont.)

$$N(X; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D * |\Sigma|}} e^{\left\{-\frac{1}{2}*(X-\mu)^T * \Sigma^{-1} * (X-\mu)\right\}}$$

Gaussian Kernel:

$\Sigma = cov[X]$ is a $D \times D$ covariance matrix:

- a symmetric matrix that stores pairwise covariance of all jointly modeled random variables
- $\Sigma_{i,j} = cov(X_i, X_j)$ - the (i, j) element

- what is the $cov(X_i, X_j)$ function? – need to reflect prior knowledge

- *Radial Basis Function* - Gaussian kernel function:

$$cov(X_i, X_j) = \exp\left(-\frac{(X_i - X_j)^2}{2\sigma^2}\right)$$

- $|X_i - X_j|$ - Euclidean Distance between X_i and X_j
- σ – is the variance;

$cov(X_i, X_j) = \exp\left(-\frac{(X_i - X_j)^2}{2}\right)$ where $\sigma = 1$ -> simplified version of RBF

$$p(\theta \mid y)$$

Gaussian Process (cont.)

Gaussian Process Model Definition:

- model of a probability distribution over possible MVN functions that fit a set of points (sample space) extracted out of a multi-dimensional vector of random variables

Key points: the prior distribution of the fitting functions is MVN

- function that updates posteriors with new observations
- probability distribution over possible functions -> in any finite interval functions are jointly Gaussian distributed
- the mean function calculated by the posterior distribution of possible functions is the function used for regression predictions

The Gaussian processes model is a distribution over functions whose shape (smoothness) is defined by Σ.

$$\boldsymbol{P(f|X) = N(f|\mu, \Sigma)}$$

$X = \{X_1, X_2, \dots, X_D\}$ – random variables vector (observed data points)

$f = \{f(X_1), f(X_2), \dots, f(X_D)\}$ – possible fitting functions

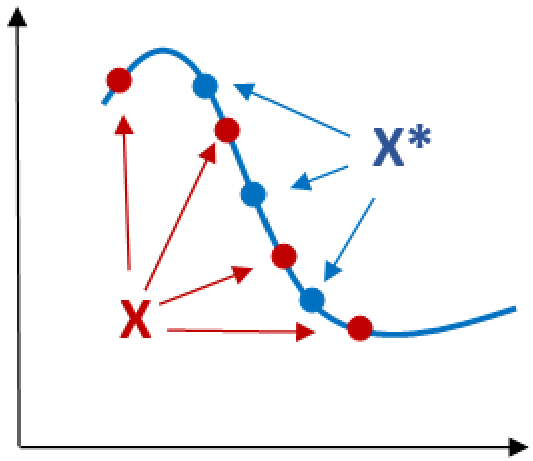
$\mu = \{\mathrm{m}(X_1), \mathrm{m}(X_2), \dots, \mathrm{m}(X_D)\}$ – μ is the mean vector, m is a mean function

$\Sigma_{i,j} = k(X_i, X_j)$ – Σ is covariance matrix, k is the Gaussian kernel function

Gaussian Process Regression

Conducting regressions by Gaussian processes:

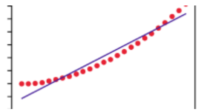
- observed data $X = \{X_1, X_2, \dots, X_D\}$ (red points)
- mean function m is estimated by the observed data points X
- m is used to make predictions at new points X^*as $\mathrm{m}(X)$



Regression Distribution Model: the regression distribution model is a conditional distribution:

$P(f^*|f, X, X^*)$

- f – the fitting functions working on X
- f^*– the fitting functions working on X^*



Summary

Bayesian Nonparametric Models – Gaussian Process Regression Model

Bayesian Nonparametric Models

Gaussian Process

- Multivariate Normal (MVN) distribution:

$$N(X;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D * |\Sigma|}} e^{\left\{-\frac{1}{2} * (X-\mu)^T * \Sigma^{-1} * (X-\mu)\right\}}$$

- Gaussian Kernel Function: *Radial Basis Function*

$$cov(X_i, X_j) = \exp\left(-\frac{(X_i - X_j)^2}{2\sigma^2}\right)$$

- Gaussian Process Model:

$$P(f|X) = N(f|\mu, \Sigma)$$

Next Lesson:

- Bayesian Nonparametric Models - Dirichlet Process Models for Clustering

Thank You!

Questions?