Gaussian Processes

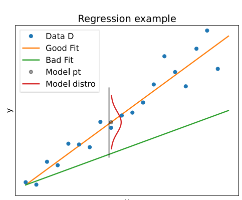
Scientific Machine Learning

Sebastian Klein

1

Motivation – Regression problems [1]

- Data $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$
- Model $f(x) = w^T x + \epsilon$
- MLE – Maximize $p(D|w)$
  $p(D|w) = \prod_{i=1}^{n} p(y_i|x_i, w)$
- MAP – Maximize $p(w|D)$
  - Bayes Theorem
  $p(w|D) = p(D|w)p(w)/p(D)$
- Assume every probability to be Gaussian -> $p(w|D) \sim \mathcal{N}(\mu, \Sigma)$

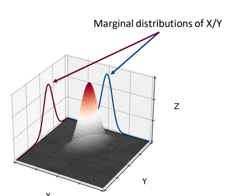


[1] K. Weinberger, Machine Learning Lecture: Gaussian Process (https://www.cs .cornell.edu/courses/cs4780/2018fa/lectures/lecturenote15.html) 5

5

Intermezzo – Marginalization [2]

- Extract partial information from $P_{X,Y} = \begin{bmatrix} X \\ Y \end{bmatrix}$
  $X \sim \mathcal{N}(\mu_X, \Sigma_X) \mid Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$
- X/Y only depending on corresponding μ_*/Σ_*
  $$p(x|X) = p_X(x) = \int_y p_{X,Y}(x, y)dy$$
  $$= \int_y p_{X|Y}(x|y)p_Y(y)dy$$
- Gets us $p(w|D) \sim \mathcal{N}(\mu, \Sigma)$



7

7

Intermezzo – Conditioning [2]

- Determine probability of Y depending on X

$$X|Y \sim \mathcal{N}(\mu_X + \Sigma_{XY}^T\, \Sigma_{YY}^{-1}\, (Y - \mu_Y),\quad \Sigma_{XX} - \Sigma_{XY}\, \Sigma_{YY}^{-1}\, \Sigma_{YX})$$
$$= \mathcal{N}(\underbrace{\Sigma_{XY}^T\, \Sigma_{YY}^{-1}\, Y}_{\text{Mean}},\quad \underbrace{\Sigma_{XX} - \Sigma_{YX}^T\, \Sigma_{YY}^{-1}\, \Sigma_{YX}}_{\text{Covariance Matrix}})$$

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{bmatrix}$$

- For $Y|X$ switch all X above with Y
- Conditioning allows to implement Bayesian Inference
  - Aka updating model with new data Y as soon as new data available

[2] C. Fonnesbeck, Fitting Gaussian Process Models in Python (https://domino. ai/blog/fitting-gaussian-process-models-python)

8

Regression as Gaussian Process (GP) [2]

- How to predict values?
  - Know $p(w|D) \sim \mathcal{N}(\mu, \Sigma)$
- What is a prediction here?
  - Sample drawn from $\mathcal{N}(\mu, \Sigma)$
- How get $\mathcal{N}(\mu, \Sigma)$?
  - Learn with training data Y
  - Predict value for test data X
  - Characteristic of $\mathcal{N}(\mu, \Sigma)$ determined solely by Σ
- How to determine μ and Σ?
  - Set $\mu = 0$
  - Simplifies conditioning to $X|Y \sim \mathcal{N}(0, \Sigma_{XX} - \Sigma_{YX}^T\, \Sigma_{YY}^{-1}\, \Sigma_{YX})$

[2] C. Fonnesbeck, Fitting Gaussian Process Models in Python (https://domino. ai/blog/fitting-gaussian-process-models-python)

9

Regression as Gaussian Process [2]

- Last result
  - $\mathcal{N}(0, \Sigma_{XX} - \Sigma_{YX}^T\, \Sigma_{YY}^{-1}\, \Sigma_{YX})$
- How to determine the Σ's?
  - $\Sigma = Cov(X, X')$ with $Cov(X, X') = k(x, x')$
  - Evaluate $k(x, x')$ pairwise
  - Covariance function $k(x, x')$ also called kernel
- Can choose any kernel that is
  - Positive semidefinite + Symmetric
  - Combinations possible

[2] C. Fonnesbeck, Fitting Gaussian Process Models in Python (https://domino. ai/blog/fitting-gaussian-process-models-python)

10

Regression as GP recipe [1,2]

- Choose appropriate kernel for data
- Determine prior distribution $\mathcal{N}(0, \Sigma_{YY})$ with training data Y
- Determine posterior distribution with conditioning X

$$\mathcal{N}(\Sigma_{XY}^T \Sigma_{YY}^{-1}\, Y, \Sigma_{XX} - \Sigma_{YX}^T\, \Sigma_{YY}^{-1}\, \Sigma_{\mathrm{YX}})$$

- If needed add noise of data σ_{Data}

$$\mathcal{N}(\Sigma_{XY}^T (\Sigma_{YY}^{-1} + \sigma^2 I) Y, \Sigma_{XX} - \Sigma_{YX}^T (\Sigma_{YY}^{-1} + \sigma^2 I) \Sigma_{\mathrm{YX}})$$

- With marginalization extract any μ_i / σ_i with $\sigma_i^2 = \Sigma_{ii}$
  - Got variance of prediction/ confidence of prediction

[1] K. Weinberger, Machine Learning Lecture: Gaussian Process (https://www.cs .cornell.edu/courses/cs4780/2018fa/lectures/lecturenote15.html)
[2] C. Fonnesbeck, Fitting Gaussian Process Models in Python (https://domino. ai/blog/fitting-gaussian-process-models-python)