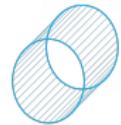




HCMUS
Viet Nam National University
Ho Chi Minh City
University of Science



Khoa Toán - Tin học
Fac. of Math. & Computer Science

CHƯƠNG 4

Phương pháp Kernel

Lưu Giang Nam

Bộ môn Ứng dụng Tin học
Khoa Toán - Tin học
Trường Đại học KHTN, ĐHQG TPHCM

09/2025

Phần 1

Giới thiệu

Giới thiệu

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

■ Mô hình tham số tuyến tính:

- Ví dụ: Hồi quy tuyến tính, Mạng nơ-ron.
- Học vector tham số w từ dữ liệu huấn luyện.
- Sau khi học, dữ liệu bị loại bỏ; dự đoán chỉ dựa trên w .

■ Phương pháp dựa trên bộ nhớ:

- Ví dụ: Nearest Neighbours, Parzen density.
- Lưu trữ toàn bộ (hoặc một phần) tập dữ liệu huấn luyện.
- Dự đoán dựa trên độ đo tương đồng giữa dữ liệu mới và dữ liệu đã lưu.
- Nhược điểm: Tốc độ dự đoán chậm khi tập dữ liệu lớn.



Hàm Kernel (Hạt nhân)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Nhiều mô hình tham số tuyến tính có thể được chuyển sang *biểu diễn đối ngẫu* (dual representation).
- Dự đoán dựa trên tổ hợp tuyến tính của hàm kernel đánh giá tại các điểm dữ liệu huấn luyện.
- Với không gian đặc trưng phi tuyến $\phi(\mathbf{x})$, hàm kernel được định nghĩa:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') \quad (1)$$

- Tính chất đối xứng:** $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$.
- Linear Kernel:** Nếu $\phi(\mathbf{x}) = \mathbf{x}$, ta có $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$.



Thủ thuật Kernel (Kernel Substitution)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở

Hạt nhân

Quá trình

Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- **Ý tưởng:** Nếu một thuật toán chỉ sử dụng tích vô hướng của các vector đầu vào, ta có thể thay thế tích vô hướng đó bằng một hàm kernel bất kỳ.
- Cho phép mở rộng các thuật toán tuyến tính sang phi tuyến.
- **Ví dụ ứng dụng:**
 - Kernel PCA (biến thể phi tuyến của PCA).
 - Support Vector Machines (SVM).
 - Kernel Fisher Discriminant.



Các loại Hàm Kernel phổ biến

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

■ **Stationary Kernels (Kernel dừng):**

- Chỉ phụ thuộc vào hiệu của các đối số.
- $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$.
- Bất biến với phép tịnh tiến trong không gian đầu vào.

■ **Homogeneous Kernels (Radial Basis Functions - RBF):**

- Chỉ phụ thuộc vào độ lớn khoảng cách (thường là Euclidean).
- $k(\mathbf{x}, \mathbf{x}') = k(\|\mathbf{x} - \mathbf{x}'\|)$.



Biểu diễn Đổi ngẫu

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Xét mô hình hồi quy tuyến tính tối thiểu hóa hàm lỗi tổng bình phương có điều chỉnh (regularized sum-of-squares):

$$J(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N (\mathbf{w}^T \phi(\mathbf{x}_n) - t_n)^2 + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w} \quad (2)$$

- Trong đó $\lambda \geq 0$.



Ma trận Thiết kế và Vector tham số \mathbf{a}

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Φ : Ma trận thiết kế, hàng thứ n là $\phi(\mathbf{x}_n)^T$.
- Vector $\mathbf{a} = (a_1, \dots, a_N)^T$ được định nghĩa:

$$a_n = -\frac{1}{\lambda}(\mathbf{w}^T \phi(\mathbf{x}_n) - t_n) \quad (3)$$

- Định nghĩa **Ma trận Gram $\mathbf{K} = \Phi \Phi^T$** .
- Đây là ma trận đối xứng kích thước $N \times N$ với các phần tử:

$$K_{nm} = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) \quad (4)$$



Giải bài toán Đồi ngẫu

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

Câu hỏi

Đặt đạo hàm của $J(\mathbf{w})$ theo \mathbf{w} bằng 0, chứng minh nghiệm \mathbf{w} là tổ hợp tuyến tính của các vector $\phi(\mathbf{x}_n)$ dạng $\mathbf{w} = \Phi^T \mathbf{a}$, từ đó tính $J(\mathbf{a})$ theo \mathbf{a} , \mathbf{t} , Φ , \mathbf{K} .

Câu hỏi

Đặt đạo hàm của $J(\mathbf{a})$ theo \mathbf{a} bằng 0, tính \mathbf{a} từ đó chứng minh:

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t} \quad (5)$$



So sánh: Primal vs. Dual Formulation

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- **Công thức gốc (Primal):** Phải nghịch đảo ma trận $M \times M$ (để tìm \mathbf{w}), với M là số chiều của không gian đặc trưng.
- **Công thức đối ngẫu (Dual):** Phải nghịch đảo ma trận $N \times N$ (để tìm \mathbf{a}), với N là số lượng điểm dữ liệu.
- **Ưu điểm của Dual:**
 - Được biểu diễn hoàn toàn thông qua hàm kernel $k(\mathbf{x}, \mathbf{x}')$.
 - Không cần làm việc trực tiếp với vector đặc trưng $\phi(\mathbf{x})$.
 - Cho phép sử dụng không gian đặc trưng có số chiều rất cao, thậm chí vô hạn.



Phần 2

Xây dựng Kernel

Xây dựng Kernel

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Để sử dụng thủ thuật thay thế kernel, ta cần xây dựng các hàm kernel hợp lệ.
- **Cách tiếp cận 1:** Chọn một ánh xạ không gian đặc trưng $\phi(\mathbf{x})$ và tìm kernel tương ứng:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') = \sum_{i=1}^M \phi_i(\mathbf{x})\phi_i(\mathbf{x}') \quad (6)$$

- **Cách tiếp cận 2:** Xây dựng trực tiếp hàm kernel $k(\mathbf{x}, \mathbf{x}')$.
- **Điều kiện hợp lệ:** Ma trận Gram \mathbf{K} (với các phần tử $k(\mathbf{x}_n, \mathbf{x}_m)$) phải là **bán xác định dương** (positive semidefinite) với mọi tập hợp $\{\mathbf{x}_n\}$.



Ví dụ: Kiểm tra tính hợp lệ của Kernel

Kiểm chứng tính hợp lệ của các Kernel sau, tìm ánh xạ đặc trưng tương ứng.

- Xét hàm kernel: $k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2$.
- Xét kernel tổng quát hơn với hằng số $c > 0$:

$$k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + c)^2 \quad (7)$$

- Giả sử ta có hai kernel hợp lệ:
 - Linear Kernel: $k_1(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \mathbf{z}$ (đặc trưng tuyến tính).
 - Quadratic Kernel: $k_2(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2$ (đặc trưng bậc 2).Kernel mới $k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) + k_2(\mathbf{x}, \mathbf{z})$ có hợp lệ không?
- Gaussian Kernel 1 chiều (giản lược):

$$k(x, z) = \exp(-(x - z)^2) \quad (8)$$

- Cho A_1, A_2 là các tập hợp con, kernel cho dữ liệu ký hiệu

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|} \quad (9)$$

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace



Kỹ thuật xây dựng Kernel mới

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

Cho trước các kernel hợp lệ $k_1(\mathbf{x}, \mathbf{x}')$ và $k_2(\mathbf{x}, \mathbf{x}')$, các kernel sau cũng hợp lệ:

- Nhân với hằng số dương: $k(\mathbf{x}, \mathbf{x}') = ck_1(\mathbf{x}, \mathbf{x}')$
- Nhân với hàm số: $k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}')$
- Đa thức hệ số không âm: $k(\mathbf{x}, \mathbf{x}') = q(k_1(\mathbf{x}, \mathbf{x}'))$
- Hàm mũ: $k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}'))$
- Tổng kernel: $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$
- Tích kernel: $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')$



Kỹ thuật xây dựng Kernel mới

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace



Các dạng xây dựng phức tạp hơn:

- Qua ánh xạ ϕ : $k(\mathbf{x}, \mathbf{x}') = k_3(\phi(\mathbf{x}), \phi(\mathbf{x}'))$
- Qua ma trận đối xứng bán xác định dương \mathbf{A} :

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{A} \mathbf{x}' \quad (10)$$

- Kết hợp các biến con (với $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_b)$):

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a) + k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (11)$$

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a)k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (12)$$

Kernel Đa thức

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Kernel cơ bản: $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}')^2$ (chứa các số hạng bậc 2).
- Kernel tổng quát hơn: $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + c)^2$ (chứa cả số hạng hằng số, bậc 1 và bậc 2).
- Tổng quát cho bậc M :

$$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + c)^M \quad (13)$$

- Ví dụ xử lý ảnh: Kernel này đại diện cho tổng có trọng số của tất cả các tích M pixel trong ảnh thứ nhất với M pixel trong ảnh thứ hai.



Gaussian Kernel

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Dạng phổ biến (còn gọi là RBF):

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (14)$$

- Chứng minh tính hợp lệ dựa trên khai triển bình phương:

$$\|\mathbf{x} - \mathbf{x}'\|^2 = \mathbf{x}^T \mathbf{x} + (\mathbf{x}')^T \mathbf{x}' - 2\mathbf{x}^T \mathbf{x}' \quad (15)$$

- Vector đặc trưng tương ứng với Gaussian Kernel có số chiều vô hạn (infinite dimensionality).
- Có thể mở rộng bằng cách thay thế $\mathbf{x}^T \mathbf{x}'$ bằng một kernel phi tuyến $\kappa(\mathbf{x}, \mathbf{x}')$.



Kernel từ Mô hình Sinh

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Kết hợp mô hình sinh (generative) và mô hình phân biệt (discriminative).
- Định nghĩa kernel dựa trên xác suất:

$$k(\mathbf{x}, \mathbf{x}') = p(\mathbf{x})p(\mathbf{x}') \quad (16)$$

- Mở rộng cho mô hình hỗn hợp (Mixture Models) với biến ẩn i :

$$k(\mathbf{x}, \mathbf{x}') = \sum_i p(\mathbf{x}|i)p(\mathbf{x}'|i)p(i) \quad (17)$$

- Hai đầu vào \mathbf{x} và \mathbf{x}' tương đồng nếu chúng có xác suất cao dưới cùng một thành phần của mô hình.



Kernel cho chuỗi (Hidden Markov Models)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở

Hạt nhân

Quá trình

Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- Xét dữ liệu là chuỗi quan sát \mathbf{X} và chuỗi trạng thái ẩn \mathbf{Z} .
- Dựa trên Hidden Markov Model (HMM), ta định nghĩa độ tương đồng:

$$k(\mathbf{X}, \mathbf{X}') = \sum_{\mathbf{Z}} p(\mathbf{X}|\mathbf{Z})p(\mathbf{X}'|\mathbf{Z})p(\mathbf{Z}) \quad (18)$$

- Cho phép so sánh các chuỗi có độ dài khác nhau.



Fisher Kernel

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Dựa trên vector tham số θ của mô hình sinh $p(\mathbf{x}|\theta)$.
- **Fisher Score:** Vector gradient trong không gian tham số:

$$\mathbf{g}(\theta, \mathbf{x}) = \nabla_{\theta} \ln p(\mathbf{x}|\theta) \quad (19)$$

- **Fisher Kernel:**

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{g}(\theta, \mathbf{x})^T \mathbf{F}^{-1} \mathbf{g}(\theta, \mathbf{x}') \quad (20)$$

- Trong đó \mathbf{F} là ma trận thông tin Fisher:

$$\mathbf{F} = \mathbb{E}_{\mathbf{x}}[\mathbf{g}(\theta, \mathbf{x}) \mathbf{g}(\theta, \mathbf{x})^T] \quad (21)$$



Sigmoidal Kernel

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại
Xấp xỉ Laplace

- Dạng hàm tanh (giống mạng nơ-ron):

$$k(\mathbf{x}, \mathbf{x}') = \tanh(a\mathbf{x}^T \mathbf{x}' + b) \quad (22)$$

- **Lưu ý:** Ma trận Gram của kernel này **không** luôn luôn bán xác định dương.
- Tuy nhiên, vẫn được sử dụng trong thực tế (ví dụ trong SVM đời đầu).
- Liên hệ lý thuyết: Mạng nơ-ron Bayesian với số lượng hàm cơ sở vô hạn hội tụ về Gaussian Process.



Phần 3

Mạng Hàm Cơ sở Hạt nhân

Radial Basis Functions network (RBF Networks)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace



- Trong các mô hình hồi quy tuyến tính, ta sử dụng tổ hợp của các hàm cơ sở cố định.
- Radial Basis Functions (RBF):** Mỗi hàm cơ sở chỉ phụ thuộc vào khoảng cách (thường là Euclidean) từ một tâm μ_j :

$$\phi_j(\mathbf{x}) = h(\|\mathbf{x} - \mu_j\|) \quad (23)$$

- Lịch sử (Nội suy hàm số chính xác):**
 - Mục tiêu: Tìm hàm $f(\mathbf{x})$ khớp chính xác với mọi giá trị mục tiêu t_n (với $n = 1, \dots, N$).
 - Đặt một hàm cơ sở tại mỗi điểm dữ liệu \mathbf{x}_n :

$$f(\mathbf{x}) = \sum_{n=1}^N w_n h(\|\mathbf{x} - \mathbf{x}_n\|) \quad (24)$$

- Nhược điểm:** Với dữ liệu nhiễu, nội suy chính xác dẫn đến quá khớp (overfitting).

Chuẩn hóa Hàm cơ sở

- Các hàm cơ sở thường được chuẩn hóa để tổng bằng 1:

$$\sum_{n=1}^N h(\mathbf{x} - \mathbf{x}_n) = 1 \quad (25)$$

- **Lợi ích:** Tránh việc dự đoán bị nhỏ hoặc chỉ phụ thuộc vào bias ở những vùng không gian đầu vào thưa thớt dữ liệu.
- Mô hình gốc có N hàm cơ sở (mỗi điểm dữ liệu là một tâm) → tốn kém khi dự đoán.
- **Giải pháp:** Sử dụng M hàm cơ sở ($M < N$).
- Các cách chọn tâm μ_i :
 - Lấy ngẫu nhiên một tập con từ dữ liệu huấn luyện.
 - **Orthogonal Least Squares:** Chọn tuân tự các điểm dữ liệu giúp giảm lỗi tổng bình phương nhiều nhất.
 - **Phân cụm (Clustering):** Dùng thuật toán K-means để tìm các tâm cụm làm tâm hàm cơ sở.
- Sau khi cố định tâm, các hệ số w được tìm bằng bình phương tối thiểu (Least Squares).

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace



Mô hình Nadaraya-Watson (Kernel Regression)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Xuất phát từ ước lượng mật độ hạt nhân (Kernel Density Estimation / Parzen Window) cho phân phối đồng thời $p(\mathbf{x}, t)$:

$$p(\mathbf{x}, t) = \frac{1}{N} \sum_{n=1}^N f(\mathbf{x} - \mathbf{x}_n, t - t_n) \quad (26)$$

- Trong đó $f(\mathbf{x}, t)$ là mật độ thành phần (component density), tập trung tại mỗi điểm dữ liệu.
- Hàm hồi quy $y(\mathbf{x})$ được định nghĩa là kỳ vọng có điều kiện:

$$y(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}] = \int_{-\infty}^{\infty} tp(t|\mathbf{x})dt = \frac{\int tp(\mathbf{x}, t)dt}{\int p(\mathbf{x}, t)dt} \quad (27)$$



Công thức Nadaraya-Watson

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace



- Giả sử các hàm mật độ thành phần có trung bình bằng 0, ta thu được:

$$y(\mathbf{x}) = \sum_{n=1}^N k(\mathbf{x}, \mathbf{x}_n) t_n \quad (28)$$

- Với hàm kernel $k(\mathbf{x}, \mathbf{x}_n)$ được định nghĩa:

$$k(\mathbf{x}, \mathbf{x}_n) = \frac{g(\mathbf{x} - \mathbf{x}_n)}{\sum_{m=1}^N g(\mathbf{x} - \mathbf{x}_m)} \quad (29)$$

- Trong đó $g(\mathbf{x}) = \int f(\mathbf{x}, t) dt$.

Tính chất và Mở rộng

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- **Trọng số cục bộ:** Kernel $k(\mathbf{x}, \mathbf{x}_n)$ gán trọng số lớn hơn cho các điểm dữ liệu \mathbf{x}_n gần \mathbf{x} .
- **Ràng buộc tổng:** $\sum_{n=1}^N k(\mathbf{x}, \mathbf{x}_n) = 1$.
- **Phân phối có điều kiện đầy đủ:**

$$p(t|\mathbf{x}) = \frac{p(t, \mathbf{x})}{p(\mathbf{x})} = \frac{\sum_{n=1}^N f(\mathbf{x} - \mathbf{x}_n, t - t_n)}{\sum_{m=1}^M g(\mathbf{x} - \mathbf{x}_m)} \quad (30)$$

- Nếu f là Gaussian, $p(t|\mathbf{x})$ là một mô hình hỗn hợp Gaussian (Gaussian Mixture).
- **Mở rộng:** Có thể sử dụng Gaussian Mixture Model (GMM) để mô hình hóa $p(t, \mathbf{x})$ trước, giúp giảm chi phí tính toán khi dự đoán.

Phần 4

Quá trình Gaussian

Quá trình Gaussian (Gaussian Processes - GP)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

■ **Cách tiếp cận truyền thống:** Mô hình tham số (như hồi quy tuyến tính $y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \phi(\mathbf{x})$).

- Xác định phân phối tiên nghiệm trên tham số \mathbf{w} .
- Suy ra phân phối trên hàm số.

■ **Cách tiếp cận GP:** Định nghĩa trực tiếp phân phối tiên nghiệm trên không gian hàm số.

- Loại bỏ mô hình tham số trung gian.
- Mặc dù không gian hàm là vô hạn, ta chỉ cần xét giá trị hàm tại tập hợp hữu hạn các điểm dữ liệu \mathbf{x}_n .

■ **Liên hệ:** Tương đương với Kriging (địa thống kê), bộ lọc Kalman, và mạng RBF.



Nhìn lại Hồi quy Tuyến tính dưới góc độ GP

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Xét mô hình: $y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$.
- Tiên nghiệm trên \mathbf{w} : $p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \mathbf{0}, \alpha^{-1} \mathbf{I})$.
- Xét vector các giá trị hàm $\mathbf{y} = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_N))^T$.
- Ta có $\mathbf{y} = \Phi \mathbf{w}$, với Φ là ma trận thiết kế.
- Vì \mathbf{y} là tổ hợp tuyến tính của các biến Gaussian, \mathbf{y} cũng phân phối Gaussian với:

$$\mathbb{E}[\mathbf{y}] = \Phi \mathbb{E}[\mathbf{w}] = \mathbf{0} \quad (31)$$

$$\text{cov}[\mathbf{y}] = \mathbb{E}[\mathbf{y}\mathbf{y}^T] = \Phi \mathbb{E}[\mathbf{w}\mathbf{w}^T] \Phi^T = \frac{1}{\alpha} \Phi \Phi^T = \mathbf{K} \quad (32)$$

- \mathbf{K} là ma trận Gram với $K_{nm} = k(\mathbf{x}_n, \mathbf{x}_m)$.



Định nghĩa Gaussian Process

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

Định nghĩa

Một Gaussian process là một phân phối xác suất trên các hàm số $y(\mathbf{x})$ sao cho tập hợp các giá trị của $y(\mathbf{x})$ tại bất kỳ tập điểm tùy ý $\mathbf{x}_1, \dots, \mathbf{x}_N$ nào cũng tuân theo phân phối Gaussian đồng thời (joint Gaussian distribution).

- Được xác định hoàn toàn bởi các thống kê bậc hai:
 - Trung bình (thường giả định bằng 0 do đối xứng): $\mathbb{E}[y(\mathbf{x})] = 0$.
 - Hiệp phương sai (xác định bởi hàm kernel):

$$\mathbb{E}[y(\mathbf{x}_n)y(\mathbf{x}_m)] = k(\mathbf{x}_n, \mathbf{x}_m) \quad (33)$$



Phần 4

Quá trình Gaussian

Mục 1: GAUSSIAN PROCESSES CHO BÀI
TOÁN HỒI QUY

Gaussian Processes cho Bài toán Hồi quy

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- Mô hình quan sát có nhiễu: $t_n = y_n + \epsilon_n$.
- Nhiều Gaussian độc lập: $p(t_n|y_n) = \mathcal{N}(t_n|y_n, \beta^{-1})$.
- Phân phối đồng thời của mục tiêu \mathbf{t} điều kiện theo \mathbf{y} :

$$p(\mathbf{t}|\mathbf{y}) = \mathcal{N}(\mathbf{t}|\mathbf{y}, \beta^{-1}\mathbf{I}_N) \quad (34)$$

- Phân phối biên của \mathbf{y} (từ định nghĩa GP):

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) \quad (35)$$



Phân phối biên của \mathbf{t}

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Tích phân \mathbf{y} để tìm phân phối biên của \mathbf{t} :

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{y})p(\mathbf{y})d\mathbf{y} = \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C}) \quad (36)$$

- Ma trận hiệp phương sai \mathbf{C} có các phần tử:

$$C(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \beta^{-1}\delta_{nm} \quad (37)$$

- Ý nghĩa: Hiệp phương sai của \mathbf{t} là tổng của hiệp phương sai tín hiệu (do kernel) và hiệp phương sai nhiễu.
- Kernel phổ biến (Exponential of quadratic form):

$$k(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 \exp\left(-\frac{\theta_1}{2}\|\mathbf{x}_n - \mathbf{x}_m\|^2\right) + \theta_2 + \theta_3 \mathbf{x}_n^T \mathbf{x}_m \quad (38)$$



Dự đoán với Gaussian Process

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- **Mục tiêu:** Dự đoán t_{N+1} cho đầu vào mới \mathbf{x}_{N+1} dựa trên dữ liệu huấn luyện \mathbf{t}_N .
- Xét phân phối đồng thời trên $N + 1$ điểm (bao gồm điểm mới):

$$p(\mathbf{t}_{N+1}) = \mathcal{N}(\mathbf{t}_{N+1} | \mathbf{0}, \mathbf{C}_{N+1}) \quad (39)$$

- Phân hoạch ma trận hiệp phương sai \mathbf{C}_{N+1} :

$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^T & c \end{pmatrix} \quad (40)$$

- Trong đó:
 - \mathbf{C}_N : Ma trận hiệp phương sai $N \times N$ của dữ liệu huấn luyện.
 - \mathbf{k} : Vector $N \times 1$ với các phần tử $k(\mathbf{x}_n, \mathbf{x}_{N+1})$.
 - $c = k(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) + \beta^{-1}$.



Kết quả Phân phối Dự đoán

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Áp dụng công thức điều kiện cho phân phối Gaussian, ta có $p(t_{N+1}|\mathbf{t})$ là phân phối Gaussian với:
- **Trung bình (Mean):**

$$m(\mathbf{x}_{N+1}) = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{t} \quad (41)$$

- **Phương sai (Variance):**

$$\sigma^2(\mathbf{x}_{N+1}) = c - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k} \quad (42)$$

- Lưu ý:
 - Trung bình là tổ hợp tuyến tính của các giá trị mục tiêu \mathbf{t} .
 - Phương sai phụ thuộc vào \mathbf{x}_{N+1} , thể hiện độ không chắc chắn của dự đoán.



Độ phức tạp tính toán: GP vs. Linear Basis

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho Phân loại
Xấp xỉ Laplace



■ Gaussian Process:

- Cần nghịch đảo ma trận \mathbf{C}_N kích thước $N \times N$.
- Độ phức tạp: $O(N^3)$.
- Dự đoán điểm mới: $O(N^2)$.
- Nhược điểm: Tốn kém khi số lượng dữ liệu N lớn.

■ Mô hình hàm cơ sở tuyến tính (Linear Basis Function):

- Cần nghịch đảo ma trận kích thước $M \times M$ (M là số hàm cơ sở).
- Độ phức tạp: $O(M^3)$.
- Hiệu quả hơn nếu $M \ll N$.

■ Ưu điểm GP: Cho phép sử dụng các hàm kernel tương ứng với số chiều vô hạn (infinite basis functions) mà không cần tính toán trực tiếp trong không gian đó.

Phần 4

Quá trình Gaussian

Mục 2: HỌC CÁC SIÊU THAM SỐ

Học các Siêu tham số

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Thay vì cố định hàm hiệp phương sai, ta sử dụng một họ hàm tham số.
- Các tham số này (gọi là **siêu tham số θ**) điều chỉnh:
 - Thang độ dài (length scale) của các tương quan.
 - Độ chính xác (precision) của nhiễu.
- **Phương pháp:** Tối đa hóa hàm log khả năng (log likelihood) $p(\mathbf{t}|\theta)$.
- Tương tự như quy trình ước lượng hợp lý cực đại loại 2 (Type 2 Maximum Likelihood) trong hồi quy tuyến tính.



Đánh giá Hàm Log Likelihood

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Hàm log likelihood cho mô hình hồi quy GP được tính dựa trên phân phối Gaussian đa biến:

$$\ln p(\mathbf{t}|\boldsymbol{\theta}) = -\frac{1}{2} \ln |\mathbf{C}_N| - \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \mathbf{t} - \frac{N}{2} \ln(2\pi) \quad (43)$$

- Để tối ưu hóa (ví dụ dùng phương pháp gradient liên hợp), ta cần đạo hàm theo các tham số θ_i :

$$\frac{\partial}{\partial \theta_i} \ln p(\mathbf{t}|\boldsymbol{\theta}) = -\frac{1}{2} \text{Tr} \left(\mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \right) + \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \mathbf{C}_N^{-1} \mathbf{t} \quad (44)$$

- **Lưu ý:** Hàm này thường không lồi (non-convex) và có thể có nhiều cực trị địa phương.

Mở rộng: Nhiều Heteroscedastic

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Mô hình chuẩn giả định phương sai nhiễu β^{-1} là hằng số.
- Phương sai nhiễu phụ thuộc vào đầu vào x .
- **Giải pháp:** Sử dụng một Gaussian Process thứ hai để mô hình hóa sự phụ thuộc của $\ln \beta(x)$ vào x .
- Điều này cho phép mô hình linh hoạt hơn đối với dữ liệu có mức độ nhiễu thay đổi theo không gian.



Tự động xác định mức độ liên quan

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Kỹ thuật mở rộng từ việc học siêu tham số: Gán một tham số thang đo riêng biệt η_i cho mỗi chiều đầu vào x_i .
- Ví dụ kernel với tự động xác định mức độ liên quan:

$$k(\mathbf{x}, \mathbf{x}') = \theta_0 \exp \left(-\frac{1}{2} \sum_{i=1}^D \eta_i (x_i - x'_i)^2 \right) \quad (45)$$

■ Cơ chế:

- Nếu η_i nhỏ \rightarrow hàm số ít nhạy cảm với biến x_i .
- Tối ưu hóa η bằng cực đại hóa khả năng (Maximum Likelihood) sẽ tự động lái $\eta_i \rightarrow 0$ đối với các biến đầu vào không quan trọng.



Ví dụ minh họa

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

■ Xét dữ liệu tổng hợp với 3 đầu vào:

- x_1 : Dự báo tốt mục tiêu t .
- x_2 : Bản sao của x_1 nhưng thêm nhiều nhiễu.
- x_3 : Sinh ngẫu nhiên, không liên quan đến t .

■ Kết quả tối ưu hóa:

- η_1 : Hội tụ về giá trị lớn.
- η_2 : Giá trị nhỏ hơn nhiều.
- η_3 : Rất nhỏ $\rightarrow x_3$ bị loại bỏ (irrelevant).

■ Kernel ARD phổ biến (Exponential-Quadratic):

$$k(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 \exp \left(-\frac{1}{2} \sum_{i=1}^D \eta_i (x_{ni} - x_{mi})^2 \right) + \theta_2 + \theta_3 \sum_{i=1}^D x_{ni} x_{mi} \quad (46)$$



Phần 4

Quá trình Gaussian

Mục 3: GAUSSIAN PROCESSES CHO PHÂN
LOẠI

Gaussian Processes cho Phân loại

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

■ **Vấn đề:** GP dự đoán giá trị trên trục số thực $(-\infty, \infty)$, nhưng phân loại cần xác suất trong $(0, 1)$.

■ **Giải pháp:**

1 Định nghĩa GP trên hàm ẩn $a(\mathbf{x})$.

2 Biến đổi $a(\mathbf{x})$ qua hàm kích hoạt phi tuyến (ví dụ logistic sigmoid):

$$y = \sigma(a(\mathbf{x})) \in (0, 1) \quad (47)$$

■ Phân phối Bernoulli cho mục tiêu $t \in \{0, 1\}$:

$$p(t|a) = \sigma(a)^t (1 - \sigma(a))^{1-t} \quad (48)$$



Phân phối Tiên nghiệm và Hiệp phương sai

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- Tiên nghiệm GP trên vector \mathbf{a}_{N+1} :

$$p(\mathbf{a}_{N+1}) = \mathcal{N}(\mathbf{a}_{N+1} | \mathbf{0}, \mathbf{C}_{N+1}) \quad (49)$$

- Ma trận hiệp phương sai \mathbf{C}_{N+1} (không có nhiều cộng tính như hồi quy, nhưng cần ν để đảm bảo xác định dương):

$$C(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \nu \delta_{nm} \quad (50)$$

- Mục tiêu là tính phân phối dự đoán $p(t_{N+1} = 1 | \mathbf{t}_N)$:

$$p(t_{N+1} = 1 | \mathbf{t}_N) = \int \sigma(a_{N+1}) p(a_{N+1} | \mathbf{t}_N) da_{N+1} \quad (51)$$



Các phương pháp Xấp xỉ

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Tích phân trên là không giải được bằng giải tích (intractable).
- Cần xấp xỉ phân phối hậu nghiệm $p(\mathbf{a}_{N+1}|\mathbf{t}_N)$ bằng một phân phối Gaussian.
- **Các kỹ thuật chính:**
 - **Variational Inference (Suy diễn biến phân):** Sử dụng cận dưới biến phân cục bộ cho hàm sigmoid.
 - **Expectation Propagation (EP):** Thường cho kết quả tốt do phân phối hậu nghiệm thực tế thường đơn mode (unimodal).
 - **Laplace Approximation:** Xấp xỉ dựa trên đạo hàm bậc hai tại điểm cực đại (mode).



Phần 4

Quá trình Gaussian

Mục 4: XẤP xỉ LAPLACE

Xấp xỉ Laplace (Laplace Approximation)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- Để đánh giá phân phối dự đoán $p(t_{N+1} = 1 | \mathbf{t}_N)$, ta cần xấp xỉ Gaussian cho phân phối hậu nghiệm $p(\mathbf{a}_{N+1} | \mathbf{t}_N)$.
- Sử dụng định lý Bayes và tính chất Gaussian có điều kiện:

$$p(\mathbf{a}_{N+1} | \mathbf{t}_N) = \int p(\mathbf{a}_{N+1} | \mathbf{a}_N) p(\mathbf{a}_N | \mathbf{t}_N) d\mathbf{a}_N \quad (52)$$

- Trong đó:
 - $p(\mathbf{a}_{N+1} | \mathbf{a}_N)$ đã biết từ kết quả hồi quy GP (công thức 6.66, 6.67).
 - $p(\mathbf{a}_N | \mathbf{t}_N)$ là phân phối hậu nghiệm cần xấp xỉ bằng phương pháp Laplace.



Xây dựng Hàm $\Psi(\mathbf{a}_N)$

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Phân phối tiên nghiệm: $p(\mathbf{a}_N) = \mathcal{N}(\mathbf{a}_N | \mathbf{0}, \mathbf{C}_N)$.
- Hàm khả năng (Likelihood) với dữ liệu độc lập:

$$p(\mathbf{t}_N | \mathbf{a}_N) = \prod_{n=1}^N e^{a_n t_n} \sigma(-a_n) \quad (53)$$

- Xấp xỉ Laplace dựa trên khai triển Taylor của log hậu nghiệm:

$$\Psi(\mathbf{a}_N) = \ln p(\mathbf{a}_N) + \ln p(\mathbf{t}_N | \mathbf{a}_N) \quad (54)$$

$$\Psi(\mathbf{a}_N) = -\frac{1}{2} \mathbf{a}_N^T \mathbf{C}_N^{-1} \mathbf{a}_N + \mathbf{t}_N^T \mathbf{a}_N - \sum_{n=1}^N \ln(1+e^{a_n}) + \text{const} \quad (55)$$



Tìm Mode của Phân phối Hậu nghiệm

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Để tìm mode \mathbf{a}_N^* , ta tính gradient của $\Psi(\mathbf{a}_N)$:

$$\nabla \Psi(\mathbf{a}_N) = \mathbf{t}_N - \boldsymbol{\sigma}_N - \mathbf{C}_N^{-1} \mathbf{a}_N \quad (56)$$

- Trong đó $\boldsymbol{\sigma}_N$ là vector chứa các phần tử $\sigma(a_n)$.
- Hessian của $\Psi(\mathbf{a}_N)$:

$$\nabla \nabla \Psi(\mathbf{a}_N) = -\mathbf{W}_N - \mathbf{C}_N^{-1} \quad (57)$$

- Với \mathbf{W}_N là ma trận đường chéo, $W_{nn} = \sigma(a_n)(1 - \sigma(a_n))$.
- Vì Hessian xác định âm, $\Psi(\mathbf{a}_N)$ lồi logarit (log convex) \rightarrow có một cực đại toàn cục duy nhất.



Thuật toán Newton-Raphson (IRLS)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Sử dụng phương pháp Newton-Raphson để tìm mode \mathbf{a}_N^* :

$$\mathbf{a}_N^{\text{new}} = \mathbf{a}_N - (\nabla \nabla \Psi)^{-1} \nabla \Psi \quad (58)$$

- Công thức cập nhật (Iterative Reweighted Least Squares):

$$\mathbf{a}_N^{\text{new}} = \mathbf{C}_N(\mathbf{I} + \mathbf{W}_N \mathbf{C}_N)^{-1}(\mathbf{t}_N - \boldsymbol{\sigma}_N + \mathbf{W}_N \mathbf{a}_N) \quad (59)$$

- Sau khi hội tụ tại \mathbf{a}_N^* , xấp xỉ hậu nghiệm là Gaussian:

$$q(\mathbf{a}_N) = \mathcal{N}(\mathbf{a}_N | \mathbf{a}_N^*, \mathbf{H}^{-1}) \quad (60)$$

- Với $\mathbf{H} = \mathbf{W}_N + \mathbf{C}_N^{-1}$.



Phân phối Dự đoán

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở
Hạt nhân

Quá trình Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho Phân loại

Xấp xỉ Laplace

- Trung bình của phân phối dự đoán cho biến tiềm ẩn a_{N+1} :

$$\mathbb{E}[a_{N+1}|\mathbf{t}_N] = \mathbf{k}^T(\mathbf{t}_N - \boldsymbol{\sigma}_N) \quad (61)$$

- Phương sai của phân phối dự đoán:

$$\mathbb{V}[a_{N+1}|\mathbf{t}_N] = c - \mathbf{k}^T(\mathbf{W}_N^{-1} + \mathbf{C}_N)^{-1}\mathbf{k} \quad (62)$$

- Cuối cùng, tính xác suất phân loại bằng cách tích phân qua hàm sigmoid (xấp xỉ):

$$p(t_{N+1} = 1|\mathbf{t}_N) \approx \int \sigma(a_{N+1})\mathcal{N}(a_{N+1}|\mathbb{E}[a], \mathbb{V}[a])da_{N+1} \quad (63)$$



Học Siêu tham số (Hyperparameter Learning)

Giới thiệu

Xây dựng Kernel

Mạng Hàm Cơ sở

Hạt nhân

Quá trình

Gaussian

Gaussian Processes cho Bài toán Hồi quy

Học các Siêu tham số

Gaussian Processes cho

Phân loại

Xấp xỉ Laplace

- Tối đa hóa hàm log likelihood biên $p(\mathbf{t}_N|\theta)$.
- Gradient của log likelihood theo tham số θ_j bao gồm hai phần:
 - 1 Phụ thuộc trực tiếp của \mathbf{C}_N vào θ_j .
 - 2 Phụ thuộc gián tiếp qua \mathbf{a}_N^* (vì mode thay đổi khi θ thay đổi).
- Công thức gradient phức tạp bao gồm các đạo hàm của \mathbf{C}_N và các số hạng liên quan đến \mathbf{W}_N .
- Sử dụng các thuật toán tối ưu hóa phi tuyến tiêu chuẩn.



Kết luận về Xấp xỉ Laplace

Giới thiệu

Xây dựng Kernel

Mảng Hàm Cơ sở
Hạt nhân

Quá trình
Gaussian

Gaussian Processes cho Bài
toán Hồi quy

Học các Siêu tham số
Gaussian Processes cho
Phân loại

Xấp xỉ Laplace

- Cung cấp một khung khổ vững chắc cho phân loại Gaussian Process.
- Đảm bảo hội tụ về nghiệm duy nhất (do tính lồi).
-
- Có thể mở rộng cho bài toán phân loại đa lớp ($K > 2$) sử dụng hàm kích hoạt softmax.
- Chi phí tính toán chủ yếu nằm ở việc nghịch đảo ma trận và quá trình lặp Newton-Raphson.



Good luck!