Dẫn nhập về suy luận Bayes (Introduction to Bayesian Inference)

VietAI Guest Lecture
2020
Nguyễn Sỹ Khánh

Mục tiêu

- Giới thiệu phương pháp suy luận Bayes - cách tư duy xác suất
- Giúp khán giả hiểu được trực giác Bayes
- Giới thiệu các ứng dụng của suy luận Bayes trong Machine Learning

Thomas Bayes (/beɪz/)

Nguồn: Wikipedia

- Nhà thống kê học, triết gia và mục sư người Anh
- Sống ở thế kỉ 18 (1701-1761)

Nhắc lại về xác suất: quy tắc cộng và quy tắc nhân

Quy tắc nhân:

$$p(a, b, c) = p(a)p(b|a)p(c|b, a)$$

Quy tắc cộng (**rời rạc**):

$$p(a) = \sum_b p(a, b) = \sum_b p(a|b)p(b)$$

Quy tắc cộng (**liên tục**):

$$p(a) = \int_b p(a, b)db$$

Mọi phương trình trong suy luận Bayes đều là hệ quả của 2 quy tắc trên.

Định lý Bayes

Tính chất của xác suất có điều kiện: xét 2 sự kiện A và B

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D.*

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Biểu thức là hàm theo *θ*. Dữ liệu *D* là duy nhất và bất biến.

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

Xác suất tiên nghiệm (prior)

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

Xác suất tiên nghiệm
(prior)

Hàm khả năng
(likelihood function)

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

Xác suất tiên nghiệm (prior)

Hàm khả năng (likelihood function)

Xác suất hậu nghiệm (posterior)

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

Xác suất tiên nghiệm
(prior)

Hàm khả năng
(likelihood function)

Xác suất hậu nghiệm
(posterior)

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Chứng cứ mô hình
(model evidence)

Suy luận Bayes

Xác suất có điều kiện: tham số mô hình *θ* và dữ liệu *D*

Xác suất tiên nghiệm (prior)

Hàm khả năng (likelihood function)

Xác suất hậu nghiệm (posterior)

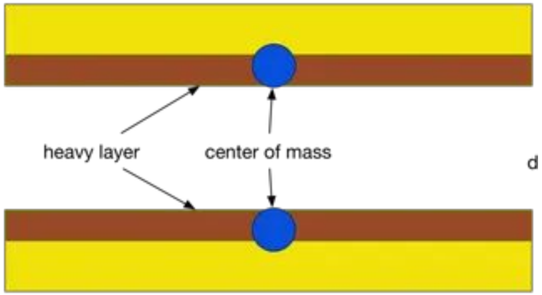
$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt.

- Nghi vấn: đồng xu bị gài (ăn gian) ⇔ xác suất ra mặt ngửa không phải 1/2
- Làm sao để kiểm định nghi vấn này?

nguồn: thinglink.com



nguồn: http://www.win-vector.com/blog/2015/04/i-still-think-you-can-manufacture-an-unfair-coin/

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt.

- Nghi vấn: đồng xu bị gài (ăn gian) ⇔ xác suất ra mặt ngửa không phải 1/2
- Làm sao để kiểm định nghi vấn này?
  ⇒ LẤY MẪU/SAMPLE

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt. Khi tung thử 3 lần, đều ra mặt ngửa.

Gọi θ là xác suất ra mặt ngửa.

Hàm khả năng:

$$p(\mathcal{D}|\theta) = \theta^3$$

⇒ θ_maximum_likelihood = 1

⇒ đồng xu có xác suất P(Ngửa) = 1

⇒ Luôn ra mặt ngửa - Có hợp lý không?

Ví dụ 1: Tung đồng xu

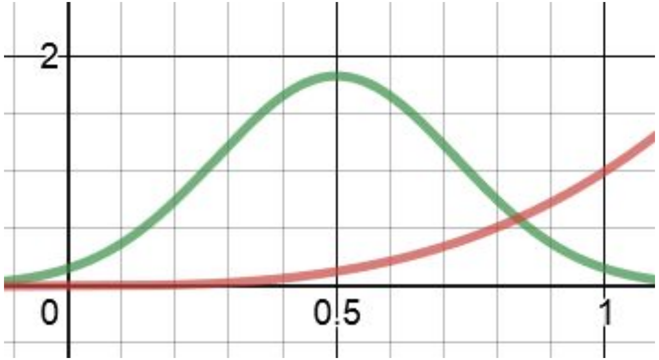
Xét một đồng xu 2 mặt. Khi tung thử 3 lần, đều ra mặt ngửa.

Tiên nghiệm: ***phân bố chuẩn*** với trung bình = 0.5

$$p(\theta) = \mathcal{N}(\theta | 0.5, \sqrt{0.05})$$

Hàm khả năng:

$$p(\mathcal{D}|\theta) = \theta^3$$



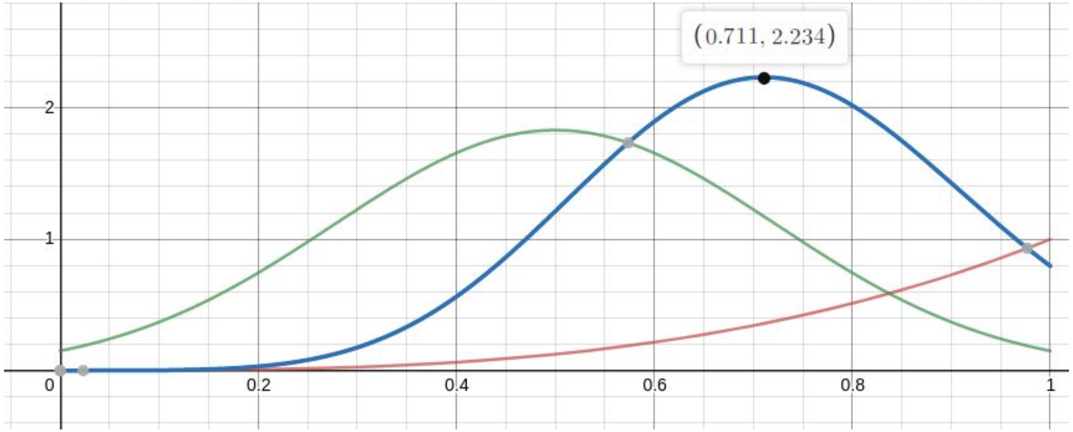
đồ thị tạo ở: https://www.desmos.com/calculator
interactive:
https://www.desmos.com/calculator/qpne2g7emb

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt. Khi tung thử 3 lần, đều ra mặt ngửa.

Hậu nghiệm:

$$p(\mathcal{D}|\theta) \propto \mathcal{N}(\theta|0.5, \sqrt{0.05})p(\mathcal{D}|\theta)$$

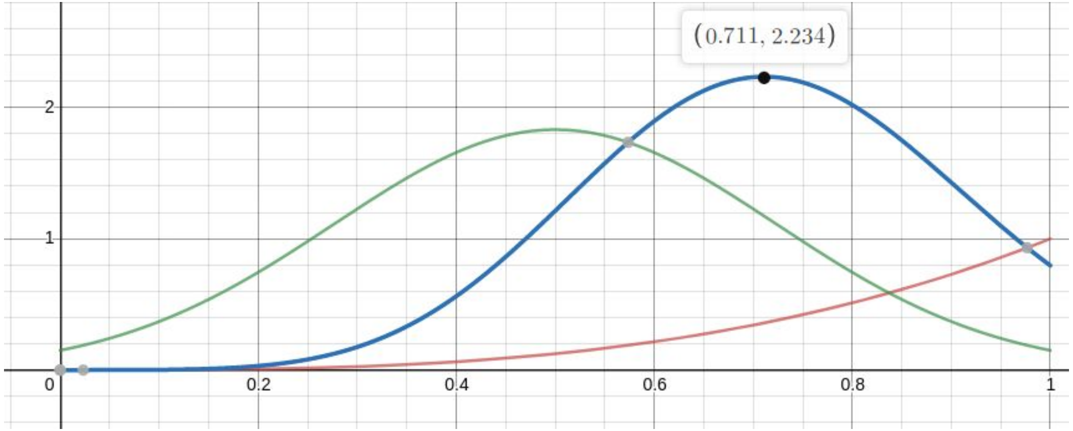


đồ thị tạo ở: https://www.desmos.com/calculator

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt. Khi tung thử 3 lần, đều ra mặt ngửa.

Hậu nghiệm:



đồ thị tạo ở: https://www.desmos.com/calculator

Ước tính giá trị của θ: cực đại hậu nghiệm (maximum a posteriori)⇒ **θ = 0.711**

Ví dụ 1: Tung đồng xu

Xét một đồng xu 2 mặt. Khi tung thử 3 lần, đều ra mặt ngửa.

Kết quả: θ là xác suất ra mặt ngửa

- Cực đại hàm khả năng: $\theta = 1$
- Suy luận Bayes: $\theta = 0.711$

Ví dụ 1: Tung đồng xu

Trường hợp khác: tung 10 lần ra 7 lần ngửa, 3 lần sấp

Likelihood: $p(\mathcal{D}|\theta) = \theta^7(1-\theta)^3$

Kết quả: θ là xác suất ra mặt ngửa

- Cực đại hàm khả năng: θ = 0.7
- Suy luận Bayes: θ = ???

Ví dụ 1: Tung đồng xu

Tiếp tục tung đồng xu thêm n lần để thu thập thêm dữ liệu.

Quá trình suy luận có thể tiếp diễn…

- tiên nghiệm mới = hậu nghiệm cũ
- Áp dụng định luật Bayes với **tiên nghiệm mới** và **hàm khả năng mới**
- Cập nhật hậu nghiệm với từng điểm/tập dữ liệu mới

Khi lượng dữ liệu tăng lên, ảnh hưởng của phân bố tiên nghiệm ban đầu giảm dần.

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu $D = \{\mathbf{X}, \mathbf{y}\}$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Cần tìm một bộ trọng số **w** để ánh xạ **x** sang **y**

$$y = \mathbf{w}^T \mathbf{x}$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Cần tìm một bộ trọng số **w** để ánh xạ **x** sang **y**

$$y = \mathbf{w}^T \mathbf{x}$$

Dùng mô hình nhiễu tín hiệu để chuyển thành bài toán xác suất

$$y = \mathbf{w}^T \mathbf{x} + \epsilon$$

$$p(\epsilon) = \mathcal{N}(0, \sigma^2)$$

$$\longrightarrow \quad p(y|\mathbf{x}, \mathbf{w}, \epsilon) = \mathcal{N}(\mathbf{w}^T \mathbf{x}, \sigma^2)$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Hàm khả năng trở thành:

$$p(\mathcal{D}|\mathbf{w}) = p(\mathbf{y}|\mathbf{w}, \mathbf{x}, \sigma) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}^T \mathbf{x}_i, \sigma^2)$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu D = {**X**, **y**}

Hàm khả năng trở thành:

$$p(\mathcal{D}|\mathbf{w}) = p(\mathbf{y}|\mathbf{w}, \mathbf{x}, \sigma) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}^T\mathbf{x}_i, \sigma^2)$$

Lấy log để tránh underflow khi tính toán xác suất - hàm **log likelihood**:

$$\ln p(\mathcal{D}|\mathbf{w}) = \sum_{i=1}^{N} \ln \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(y_i - \mathbf{w}^T\mathbf{x})^2}{2\sigma^2})$$

$$\ln p(\mathcal{D}|\mathbf{w}) = N \ln \frac{1}{\sigma} - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x})^2$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Hàm khả năng trở thành:

$$p(\mathcal{D}|\mathbf{w}) = p(\mathbf{y}|\mathbf{w}, \mathbf{x}, \sigma) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}^T\mathbf{x}_i, \sigma^2)$$

Lấy log để tránh underflow khi tính toán xác suất - hàm **log likelihood**:

$$\ln p(\mathcal{D}|\mathbf{w}) = \sum_{i=1}^{N} \ln \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(y_i - \mathbf{w}^T\mathbf{x})^2}{2\sigma^2})$$

$$\ln p(\mathcal{D}|\mathbf{w}) = N \ln \frac{1}{\sigma} - \frac{N}{2} \ln(2\pi) \boxed{- \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mathbf{w}^T\mathbf{x})^2}$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Hàm khả năng cao ⇔ xác suất cao mô hình khớp với dữ liệu

Để cực đại hóa hàm khả năng:

$$\ln p(\mathcal{D}|\mathbf{w}) = N \ln \frac{1}{\sigma} - \frac{N}{2}\ln(2\pi) \boxed{- \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x})^2}$$

cần cực tiểu hóa:

$$\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x})^2$$

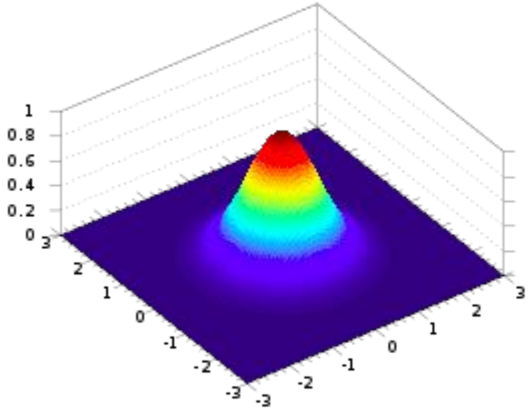
Mô hình nhiễu với phân phối chuẩn/Gaussian ⇔ hàm loss trong linear regression

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu D = {**X**, **y**}

Chọn tiên nghiệm:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|0, \mathbf{I}\alpha^{-1})$$



tiên nghiệm cho
w ∈ R^2

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

tiên nghiệm:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{I}\alpha^{-1})$$

Hậu nghiệm và log hậu nghiệm:

$$p(\mathbf{w}|\mathcal{D}) \propto \mathcal{N}(\mathbf{0}, \mathbf{I}\alpha^{-1}) * \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}^T\mathbf{x}_i, \sigma^2)$$

$$\log p(\mathbf{w}|\mathcal{D}) = -\frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x})^2 + \text{const}$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

tiên nghiệm:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{I}\alpha^{-1})$$

Hậu nghiệm và log hậu nghiệm:

$$p(\mathbf{w}|\mathcal{D}) \propto \mathcal{N}(\mathbf{0}, \mathbf{I}\alpha^{-1}) * \prod_{i=1}^{N} \mathcal{N}(\mathbf{w}^T\mathbf{x}_i, \sigma^2)$$

L2 Regulariser

$$\log p(\mathbf{w}|\mathcal{D}) = \boxed{-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x})^2 + \text{const}$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Yêu cầu dự đoán y* từ một điểm dữ liệu **x*** mới.

Phân phối dự đoán:

$$p(y^\star|\mathcal{D}, \mathbf{x}^\star) = \int p(y^\star|\mathbf{w}, \mathcal{D}, \mathbf{x}^\star)p(\mathbf{w})d\mathbf{w}$$

Ví dụ 2: Linear regression

Bài toán hồi quy tuyến tính cho một tập dữ liệu *D* = {**X**, **y**}

Yêu cầu dự đoán y* từ một điểm dữ liệu **x*** mới.

Phân phối dự đoán:

$$p(y^\star|\mathcal{D},\mathbf{x}^\star) = \boxed{\int p(y^\star|\mathbf{w},\mathcal{D},\mathbf{x}^\star)p(\mathbf{w})d\mathbf{w}}$$

tích phân phức tạp (intractable)

Ví dụ 3: Logistic regression

Mô hình: xét các điểm dữ liệu (x,y) với y = 1 hoặc 0

- Xác suất y thuộc lớp 1/0

$$p(y^{(i)} = 1|x^{(i)}) = \frac{1}{1 + e^{-w^T x^{(i)}}} = \sigma(w^T x^{(i)}) = h(x^{(i)})$$

$$p(y^{(i)} = 0|x^{(i)}) = 1 - h(x^{(i)})$$

Ví dụ 3: Logistic regression

Mô hình: xét các điểm dữ liệu (x,y) với y = 1 hoặc 0

- Likelihood của tập dữ liệu:

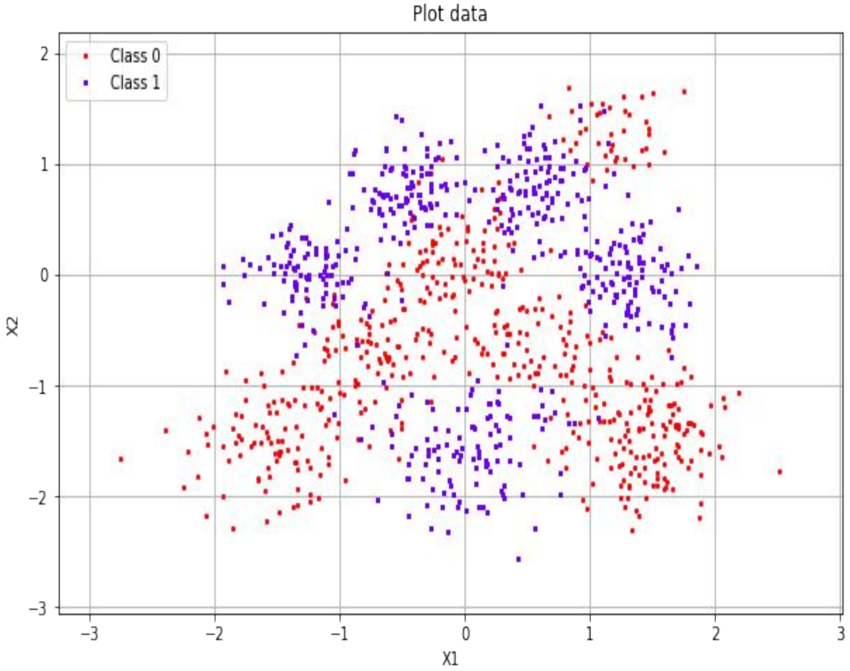
$$p(D|w) = \prod_{i=1}^{N} p(y^{(i)}|x^{(i)}) = \prod_{i=1}^{N} h(x^{(i)})^{y^{(i)}}(1 - h(x^{(i)}))^{1-y^{(i)}}$$

- Log likelihood của tập dữ liệu (= - cost function):

$$\log p(D|w) = \sum_{i=1}^{N}[y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))]$$

Ví dụ 3: Logistic regression

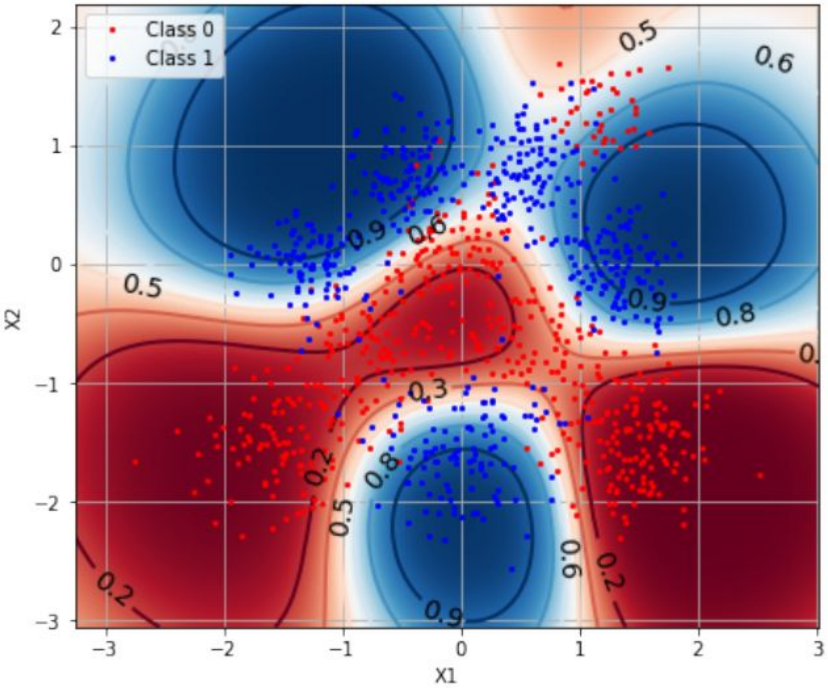
Dữ liệu:



Ví dụ 3: Logistic regression

Cực đại hàm log khả năng:

$w^* = \text{argmax} \log p(D|w)$



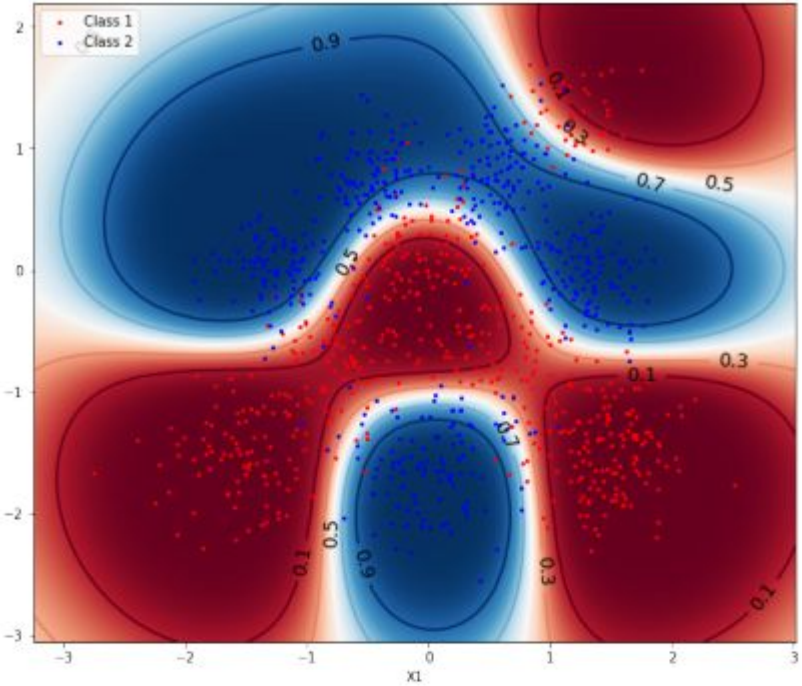
Ví dụ 3: Logistic regression

Cực đại hàm log hậu nghiệm (maximum a posteriori):

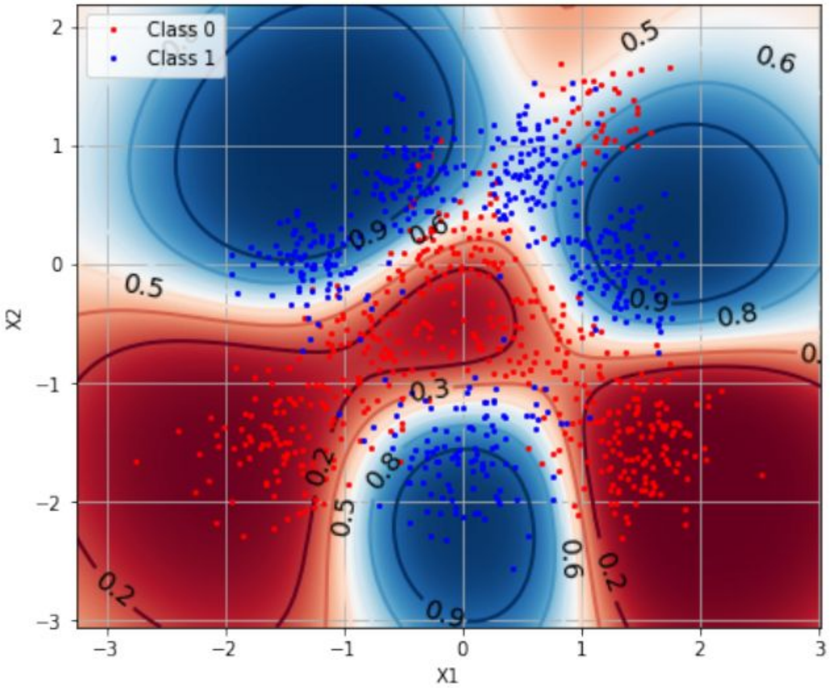
w^* = argmax log p(w|D)

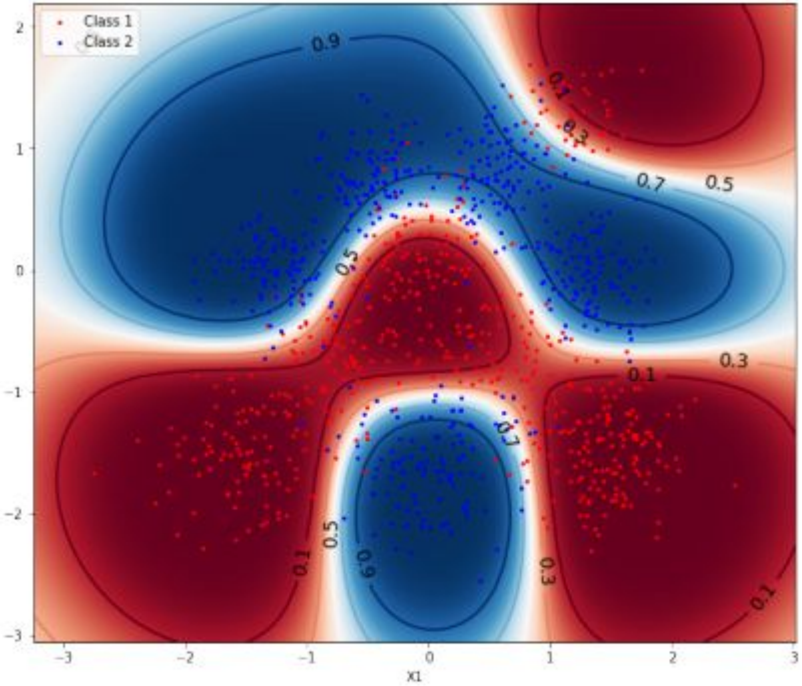
= argmax log [p(D|w)p(w)]

⇒ regularisation



Ví dụ 3: Logistic regression



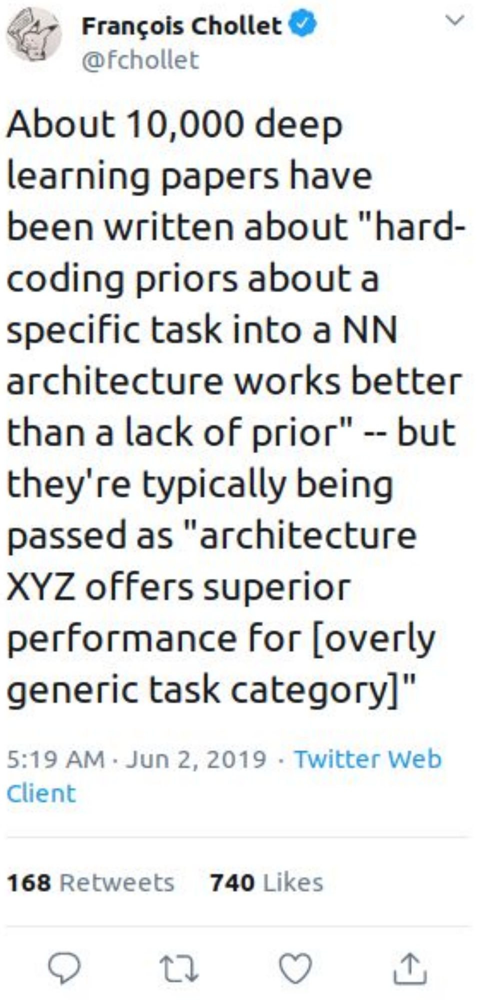


Đối chiếu

Suy luận Bayes (Bayesian approach)	Suy luận Tần suất (Frequentist approach)
Xác suất = cấp độ niềm tin (degree of belief) VD: bác sĩ chẩn đoán 60% mắc bệnh	Xác suất = tần suất xảy ra của sự kiện
Xem dữ liệu là bằng chứng duy nhất và **hữu hạn**	Xem dữ liệu như các mẫu lấy từ *vũ-trụ-dữ-liệu* **vô hạn**

Các ưu điểm

- Khi xây dựng mô hình, các **kiến thức chuyên ngành/thông tin phụ trợ** có thể được đưa vào mô hình dưới dạng **tiên nghiệm**
  - **tiên nghiệm** có thể được biểu diễn qua **kiến trúc mạng**
  - https://twitter.com/fchollet/status/1134947629576597504
  - VD: CNN thiết lập prior về tính chất của ảnh 2D bằng cách sử dụng tầng convolution
- Kết quả cuối là phân phối xác suất ⇒ chứa thông tin về rủi ro/bất định
  - Quan trọng khi ra quyết định dựa theo mô hình (trong y tế, giao thông, pháp luật, v.v.)

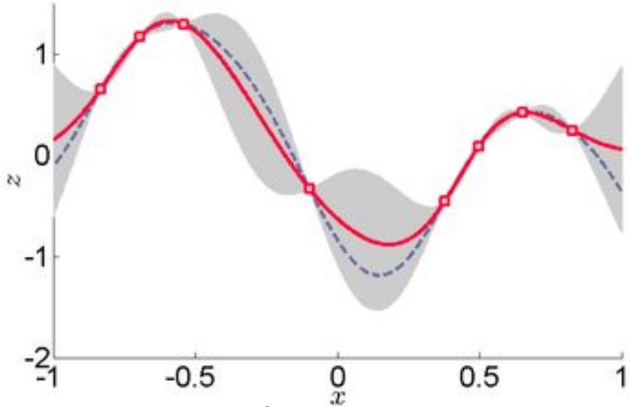


Các khuyết điểm

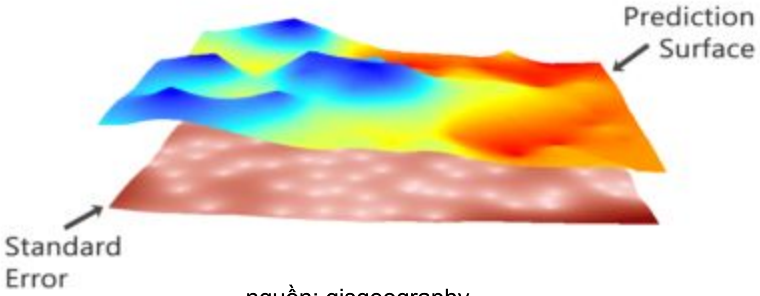
- Mathematical difficulty: phải xử lý các phép tích phân phức tạp.
  Giải pháp:
  - Các phương pháp ước lượng
  - Các phép thử: *Markov Chain Monte Carlo (MCMC), Gibbs, importance, Metropolis*, v.v
  - Các phương pháp *variational*
- Cần chọn tiên nghiệm thích hợp

Ứng dụng

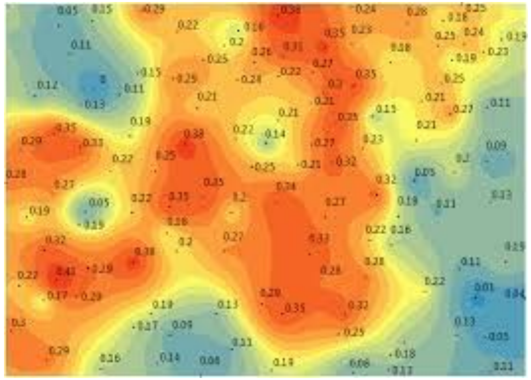
- Địa lý thống kê



nguồn: Wikipedia



nguồn: gisgeography



nguồn: gisgeography

Ứng dụng

- Địa lý thống kê
- Khí hậu học

Prediction of Himalayan precipitation from short and fragmented records

Lead Supervisor: Daniel Bannister, British Antarctic Survey

NERC criteria apply

Atmosphere, Ice, and Climate group

Rainfall in the Himalayas provides freshwater to over a billion people. However, highly localised extreme events can cause flash floods that devastate vulnerable local communities. Establishing the frequency of extreme events across the Himalayas is highly challenging, given the scarcity, reliability, and uneven distribution of rainfall observations. In this project, we aim to use a novel technique, called "patched kriging", to exploit all the information available from a network of short and fragmented rainfall measurements to produce a spatially homogenous dataset for the northwestern Himalayas. The result will be a reconstruction of annual maximum values of precipitation in ungauged/unobserved areas, demonstrating the reliability of this method for understanding the hydrometeorology of rainfall extremes in the Himalayas.

Links to relevant supporting information: Libertino, A. et al. (2018). Regional-scale analysis of extreme precipitation from short and fragmented records. Advances in Water Resources, 112, 147-159.

Duties of the student:

- Learn about rainfall observations and extreme precipitation events in the Himalayas
- Learn about statistical interpolation and machine learning methods for dealing with unevenand fragmented rainfall records
- Employ these tools to provide extreme rainfall estimates in ungauged/unobserved areas ofthe Himalayas
- Prepare a short report

Suggested timetable:

- Week 1: Welcome to BAS; literature review
- Week 2: Learn how to access, analyse, and visualise precipitation data
- Week 3: Learn about interpolation techniques and machine learning methods
- Weeks 4-8: Write and implement code to apply "patched kriging" to precipitation dataset
- Week 9-10: Write up summary; give short presentation to small group at BAS

nguồn: CUED UROP

Ứng dụng

- Địa lý thống kê
- Khí hậu học
- Điều khiển - Tự động hóa
    - Kalman filter, particle filter

Ứng dụng

- Địa lý thống kê
- Khí hậu học
- Điều khiển - Tự động hóa
  - Kalman filter, particle filter
- Khoa học cơ bản:
  - Sinh học tiên hóa: phát sinh chủng loại
    https://en.wikipedia.org/wiki/Bayesian_inference_in_phylogeny
  - Hóa sinh phân tử: BATMAN (**BA**yesian **T**ool for **M**ethylation **AN**alysis)
    https://en.wikipedia.org/wiki/Bayesian_tool_for_methylation_analysis
  - Phương pháp khoa học (scientific method)

Ứng dụng

- Địa lý thống kê
- Khí hậu học
- Điều khiển - Tự động hóa
  - Kalman filter, particle filter
- Khoa học cơ bản:
  - Sinh học tiên hóa: phát sinh chủng loại
  - Hóa sinh phân tử: BATMAN (**BA**yesian **T**ool for **M**ethylation **AN**alysis)
  - Phương pháp khoa học (scientific method)
- Ra quyết định/Decision making:
  - Y tế
  - Luật pháp
  - Chính sách
  - v.v

Ứng dụng: học máy xác suất

Naive Bayes Classifier

Bayesian Network

Dirichlet Latent allocation (LDA)

Variational Inference

Variational Auto-Encoder

$$p(C_k \mid x_1, \ldots, x_n)$$

$$p(C_k \mid \mathbf{x}) = \frac{\boxed{p(C_k)\, p(\mathbf{x} \mid C_k)}}{p(\mathbf{x})}$$

$$\boxed{p(C_k, x_1, \ldots, x_n)}$$

$$\begin{aligned}
&= p(x_1, \ldots, x_n, C_k) \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2, \ldots, x_n, C_k) \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k)\, p(x_3, \ldots, x_n, C_k) \\
&= \ldots \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k)\, p(x_2 \mid x_3, \ldots, x_n, C_k) \ldots p(x_{n-1} \mid x_n, C_k)\, p(x_n \mid C_k)\, p(C_k)
\end{aligned}$$

Giả định "naive" về tính độc lập: $p(x_i \mid x_{i+1}, \ldots, x_n, C_k) = p(x_i \mid C_k)$

$$\begin{aligned}
p(C_k \mid x_1, \ldots, x_n) &\propto p(C_k, x_1, \ldots, x_n) \\
&= p(C_k)\, p(x_1 \mid C_k)\, p(x_2 \mid C_k)\, p(x_3 \mid C_k) \cdots \\
&= p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)\,,
\end{aligned}$$

Ứng dụng: học máy xác suất

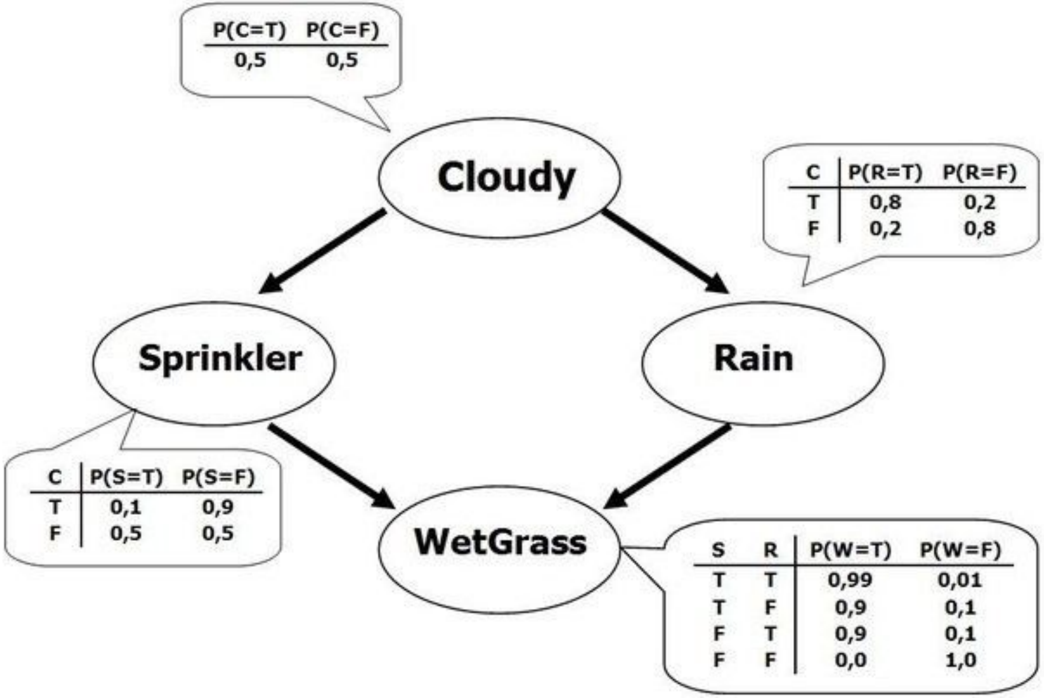
Naive Bayes Classifier

Bayesian Network

Dirichlet Latent allocation (LDA)

Variational Inference

Variational Auto-Encoder



nguồn: towardsdatascience.com

Ứng dụng: học máy xác suất

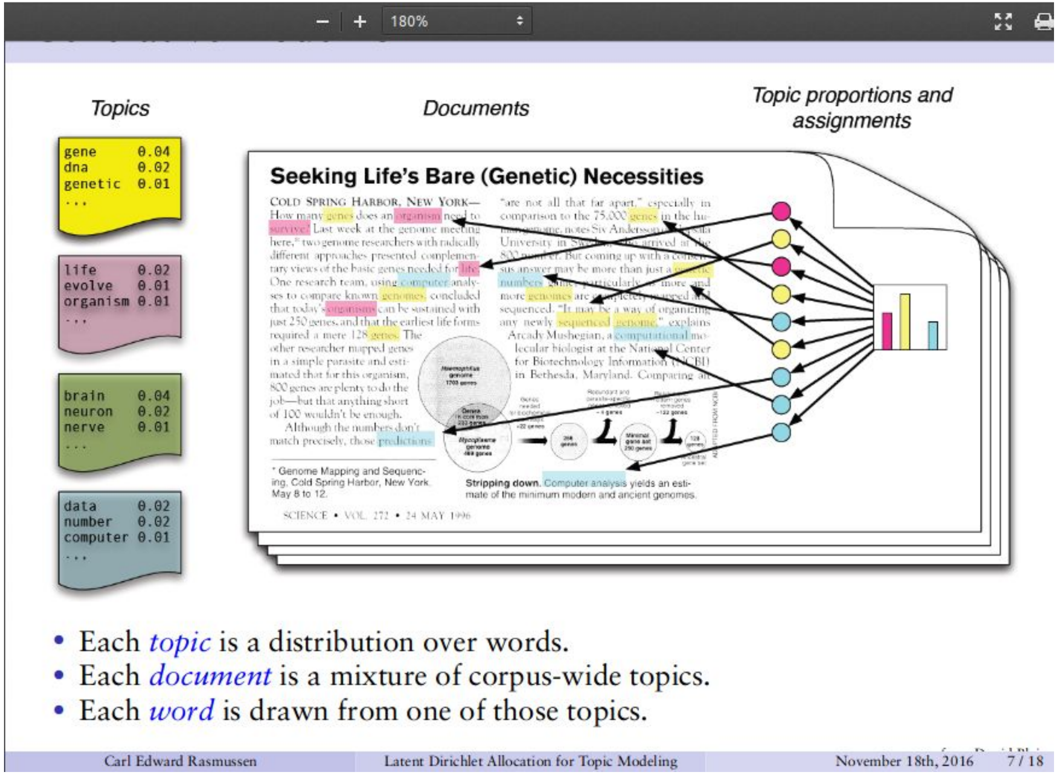
Naive Bayes Classifier

Bayesian Network

Dirichlet Latent allocation (LDA)

Variational Inference

Variational Auto-Encoder



nguồn: CUED 4F13

Ứng dụng: học máy xác suất

Naive Bayes Classifier

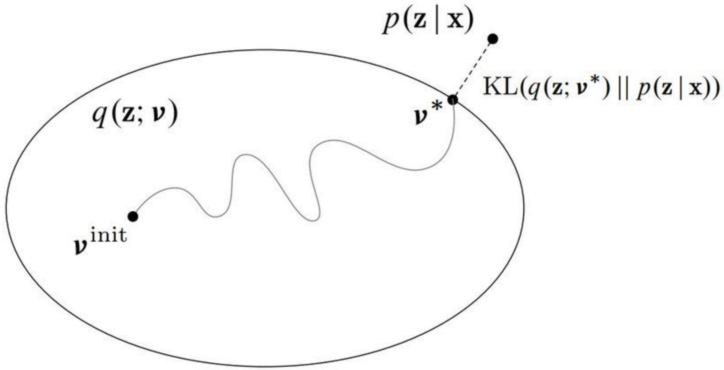
Bayesian Network

Dirichlet Latent allocation (LDA)

Variational Inference

Variational Auto-Encoder

Variational Inference



- VI turns **inference into optimization.**
- Posit a **variational family** of distributions over the latent variables,

$$q(\mathbf{z}; \boldsymbol{\nu})$$

- Fit the **variational parameters** $\boldsymbol{\nu}$ to be close (in KL) to the exact posterior.
  (There are alternative divergences, which connect to algorithms like EP, BP, and others.)

nguồn: https://media.nips.cc/Conferences/2016/Slides/6199-Slides.pdf

Ứng dụng: học máy xác suất

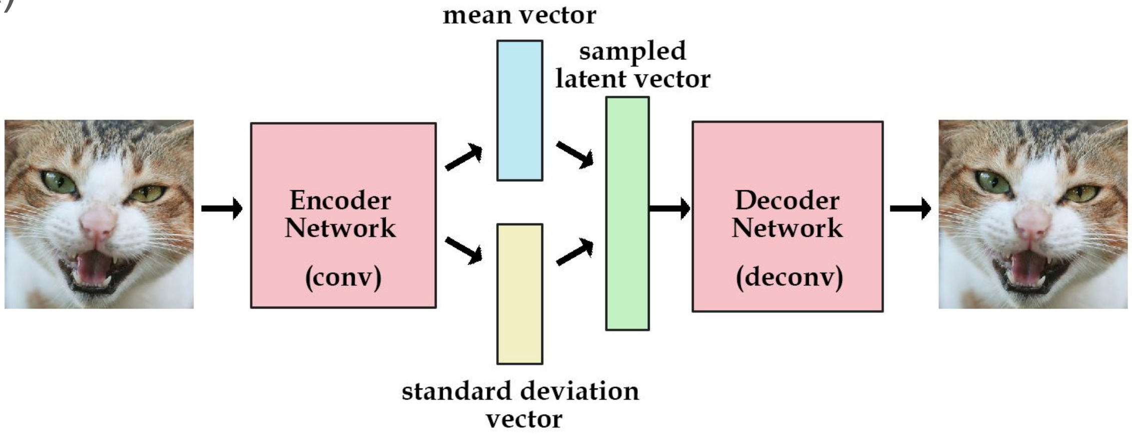
Naive Bayes Classifier

Bayesian Network

Dirichlet Latent allocation (LDA)

Variational Inference

Variational Auto-Encoder



nguồn: kvfrans.com

Các nguồn tham khảo

Pattern Recognition and Machine Learning, by Christopher Bishop

Information Theory, Inference, and Learning Algorithms, by David J.C. MacKay

https://ben-lambert.com/bayesian-lectures/

http://kvfrans.com/variational-autoencoders-explained/

https://www.cs.jhu.edu/~jason/tutorials/variational.html

https://twitter.com/