



Naive Bayes

*E-mail: pima.vn@gmail.com

Mô tả dự án

Được nghiên cứu rộng rãi từ thập niên 50, Naive Bayes là một phương pháp phân loại bắt nguồn từ lý thuyết Bayes và điều giả sử rằng các sự kiện trong điểm dữ liệu đều phân biệt. Phương pháp này nổi tiếng về việc có mô hình đơn giản và tương đối hiệu quả với những vấn đề mà ta không có quá nhiều dữ liệu.

Trong trại hè PIMA, các bạn sẽ được học về Naive Bayes và cách áp dụng phương pháp này cho mục đích phân loại.

Yêu Cầu

Trình bày mô hình Naive Bayes và, ngoài lý thuyết căn bản, hãy trả lời một số vấn đề nâng cao sau.

- (1) Trong những trường hợp cụ thể nào thì Naive Bayes thật sự hiệu quả?
- (2) Áp dụng phương pháp classification cho những class liên tục thay vì rời rạc (ví dụ như xác định một người thuộc khoảng tuổi nào hay một cổ vật là thuộc thế kỉ nào)
- (3) Tìm cách xác định xác suất prior hợp lý khi data ban đầu (training data) quá lệch về một hay một vài nhóm nào đó.
- (4) Tìm cách cập nhật xác suất prior khi ta có thêm data trong quá trình thử nghiệm (testing).
- (5) Làm sao để chọn những sự kiện phù hợp trong điểm dữ liệu để chúng đủ phân biệt để trong ra kết quả có độ chính xác cao?
- (6) Ứng dụng Naive Bayes vào một vấn đề thực tế và lập trình ra được kết quả cụ thể và tương đối hiệu quả.

Một số từ khóa: Bayesian Inference, Normal Distribution, Maximum-likelihood, Cross-validation, Correlation, Scikit-Learn.

Lưu ý: Đề tài này đòi hỏi các bạn phải có áp dụng nâng cao hơn so với những gì được đã được dạy trong bài giảng.

Tham Khảo

- [1] Các bài giảng PiMA 2018.
- [2] Rennie; Shih; Teevan; Karger (2003). *Tackling the poor assumptions of Naive Bayes classifiers*.
- [3] Murtaza Haider (2017). Cognitive Class. *Statistics 101 - Correlation*.
- [4] Irina Rish (2001). *An empirical study of the naive Bayes classifier*.