MLPR review lecture

Mainly an opportunity to ask questions

Answer exam questions

Topics list

Iain Murray
`http://iainmurray.net/`

The exam

Answer Q1, and then one of Q2 or Q3.

2hrs, 25 marks a question $\Rightarrow \approx 2$ mins/mark + reading time

Get the past papers, practice doing to time.

Read Q2 and Q3 before picking one

Reading the question:
— Identified every question asked in each part?
— . . . actual and specific questions asked?
— Can skip parts, do Q out of order, get easy marks first

Reminder: First-class / distinction level: 70%. Pass: 40% UG, 50% MSc

Exam questions

2014 q1a, 4 marks

Consider a binary classification problem. For each training instance i, where $i \in \{1, 2, \ldots N\}$, the data contains two real-valued features, $x_{i1} \in \mathbb{R}$, and $x_{i2} \in \mathbb{R}$, and a binary class label $y_i \in \{0, 1\}$. You have reason to believe that given the class label, each of the feature values is distributed approximately according to a Gaussian distribution.

Describe a naive Bayes model for the situation above. Clearly indicate which are the parameters of the model. Give the equation for the log likelihood of this model. Your answer should be specific to this model.

q1b, 3 marks

What assumption does the naive Bayes classifier make about the classification problem? If you thought that this assumption was unrealistic, how would you relax this assumption?

Exam questions

2014 q1c, i) 2 marks, ii) 4 marks

You are working for an Internet advertising agency. One particular advertisement has been displayed to three different users, and each time you have recorded a binary value $x_i \in \{0, 1\}$, for $i \in \{1, 2, 3\}$, indicating whether the user clicked on the advertisement. Let π indicate the probability that a user, sampled at random from the entire population of users, would click on the ad.

i. What is the maximum likelihood estimator of π?

ii. Before you showed the advertisement to any users, you believed based on your experience that this advertisement is either highly effective (π very near 0.7) or highly ineffective (π very near 0.2). Describe how you could incorporate this information into your analysis.

Exam questions

2014 q1c, iii) 3 marks

Compare the methods that you suggested in Question 1(c)i and Question 1(c)ii. In what situations might the method from Question 1(c)i be a better choice? In what situations might the method from Question 1(c)ii be a better choice? Are there situations in which both methods will perform similarly? Explain your answers.

q1d, 3 marks

Briefly compare and contrast logistic regression and neural networks. Describe one aspect in which the two methods are similar. Describe one aspect in which the two methods are different.

Exam questions

2011 q1d, 3 marks

Suppose you wish to predict the apartment prices in Manhattan (which has a grid street structure) using information about the geographic location and size of the flat. You decide to use regression with a linear parameter model having geographically local features. Briefly discuss what form of features you would choose.

And now stop looking at exam questions until you've done some revision. But think of possible questions as you revise. . .

Topics

Not exhaustive! Make your own lists.

Standard probability distributions:

Gaussian: $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$, $\quad \mathcal{N}(\mathbf{x}; \mu, \Sigma) = |2\pi\Sigma|^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu)}$

Gamma, Beta, Bernoulli, Binomial, deterministic $\delta(f - \mathbf{w}^\top \mathbf{x})$

Need for assumptions, building models

Class conditional models, and Naive Bayes

Regression models: deterministic mapping + noise

Logistic classifiers: sigmoid output $\rightarrow$ probability

Basis functions. Can learn them (neural nets). Or ∞, GPs.

Mixtures. Preprocessing, PCA.

Topics

Fitting models

(Penalized) Maximum Likelihood (ML or MAP).

Naive Bayes: fit easy parts, then Bayes rule to predict

Logistic regression, neural nets: gradient descent

Mixtures: EM (or gradient descent if know how)

Model selection

Fitting more complicated models gives higher likelihood

Cross-validate model choice, or. . .
integrate parameters to get marginal/model likelihood

also visualize what you're doing and perform checks

Topics

Bayesian prediction

$P(y \,|\, \mathbf{x}, \mathrm{data}) \neq P(y \,|\, \mathbf{x}, \theta_{\mathrm{MAP}})$

$P(y \,|\, \mathbf{x}, \mathrm{data}) = \int P(y \,|\, \mathbf{x}, \theta)\, p(\theta \,|\, \mathrm{data})\; \mathrm{d}\theta$

or $P(\mathbf{x}^{(N+1)} \,|\, \mathrm{data}) = \int P(\mathbf{x}^{(N+1)} \,|\, \theta)\, p(\theta \,|\, \mathrm{data})\; \mathrm{d}\theta$ (for unsupervised)

Linear regression, Gaussian processes, conjugate models:
Identify params of posterior and look up answer to integral

Approximate Inference

Monte Carlo, and MCMC to get samples from $p(\theta \,|\, \mathrm{data})$

Laplace approx., KL objectives for variational methods