Building Probabilistic Graphical Models with Python

Solve machine learning problems using probabilistic graphical models implemented in Python with real-world applications

Kiran R Karkera
# Table of Contents

## Preface 1

### Chapter 1: Probability 5

- The theory of probability 5
- Goals of probabilistic inference 8
- Conditional probability 9
- The chain rule 9
- The Bayes rule 9
- Interpretations of probability 11
- Random variables 13
- Marginal distribution 13
- Joint distribution 14
- Independence 14
- Conditional independence 15
- Types of queries 16
  - Probability queries 16
  - MAP queries 16
- Summary 18

### Chapter 2: Directed Graphical Models 19

- Graph terminology 19
  - Python digression 20
- Independence and independent parameters 20
- The Bayes network 23
  - The chain rule 24
- Reasoning patterns 24
  - Causal reasoning 25
  - Evidential reasoning 27
  - Inter-causal reasoning 27
Table of Contents

D-separation 29
  The D-separation example 31
    Blocking and unblocking a V-structure 33
Factorization and I-maps 34
The Naive Bayes model 34
  The Naive Bayes example 36
Summary 37

Chapter 3: Undirected Graphical Models 39
  Pairwise Markov networks 39
  The Gibbs distribution 41
  An induced Markov network 43
  Factorization 43
  Flow of influence 44
  Active trail and separation 45
  Structured prediction 45
    Problem of correlated features 46
    The CRF representation 46
    The CRF example 47
  The factorization-independence tango 48
  Summary 49

Chapter 4: Structure Learning 51
  The structure learning landscape 52
  Constraint-based structure learning 52
    Part I 52
    Part II 53
    Part III 54
    Summary of constraint-based approaches 60
  Score-based learning 60
    The likelihood score 61
    The Bayesian information criterion score 62
    The Bayesian score 63
    Summary of score-based learning 68
  Summary 68

Chapter 5: Parameter Learning 69
  The likelihood function 71
  Parameter learning example using MLE 72
  MLE for Bayesian networks 74
  Bayesian parameter learning example using MLE 75
  Data fragmentation 77
<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effects of data fragmentation on parameter estimation</strong></td>
<td>77</td>
</tr>
<tr>
<td><strong>Bayesian parameter estimation</strong></td>
<td>79</td>
</tr>
<tr>
<td>An example of Bayesian methods for parameter learning</td>
<td>80</td>
</tr>
<tr>
<td><strong>Bayesian estimation for the Bayesian network</strong></td>
<td>85</td>
</tr>
<tr>
<td>Example of Bayesian estimation</td>
<td>85</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>91</td>
</tr>
<tr>
<td><strong>Chapter 6: Exact Inference Using Graphical Models</strong></td>
<td>93</td>
</tr>
<tr>
<td><strong>Complexity of inference</strong></td>
<td>93</td>
</tr>
<tr>
<td>Real-world issues</td>
<td>94</td>
</tr>
<tr>
<td><strong>Using the Variable Elimination algorithm</strong></td>
<td>94</td>
</tr>
<tr>
<td>Marginalizing factors that are not relevant</td>
<td>97</td>
</tr>
<tr>
<td>Factor reduction to filter evidence</td>
<td>98</td>
</tr>
<tr>
<td>Shortcomings of the brute-force approach</td>
<td>100</td>
</tr>
<tr>
<td>Using the Variable Elimination approach</td>
<td>100</td>
</tr>
<tr>
<td><strong>Complexity of Variable Elimination</strong></td>
<td>106</td>
</tr>
<tr>
<td><strong>Graph perspective</strong></td>
<td>107</td>
</tr>
<tr>
<td>Learning the induced width from the graph structure</td>
<td>109</td>
</tr>
<tr>
<td><strong>The tree algorithm</strong></td>
<td>110</td>
</tr>
<tr>
<td>The four stages of the junction tree algorithm</td>
<td>111</td>
</tr>
<tr>
<td>Using the junction tree algorithm for inference</td>
<td>112</td>
</tr>
<tr>
<td>Stage 1.1 – moralization</td>
<td>113</td>
</tr>
<tr>
<td>Stage 1.2 – triangulation</td>
<td>114</td>
</tr>
<tr>
<td>Stage 1.3 – building the join tree</td>
<td>114</td>
</tr>
<tr>
<td>Stage 2 – initializing potentials</td>
<td>115</td>
</tr>
<tr>
<td>Stage 3 – message passing</td>
<td>115</td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>119</td>
</tr>
<tr>
<td><strong>Chapter 7: Approximate Inference Methods</strong></td>
<td>121</td>
</tr>
<tr>
<td><strong>The optimization perspective</strong></td>
<td>121</td>
</tr>
<tr>
<td>Belief propagation in general graphs</td>
<td>122</td>
</tr>
<tr>
<td>Creating a cluster graph to run LBP</td>
<td>123</td>
</tr>
<tr>
<td>Message passing in LBP</td>
<td>124</td>
</tr>
<tr>
<td><strong>Steps in the LBP algorithm</strong></td>
<td>125</td>
</tr>
<tr>
<td>Improving the convergence of LBP</td>
<td>126</td>
</tr>
<tr>
<td>Applying LBP to segment an image</td>
<td>126</td>
</tr>
<tr>
<td>Understanding energy-based models</td>
<td>128</td>
</tr>
<tr>
<td>Visualizing unary and pairwise factors on a 3 x 3 grid</td>
<td>129</td>
</tr>
<tr>
<td>Creating a model for image segmentation</td>
<td>130</td>
</tr>
<tr>
<td>Applications of LBP</td>
<td>135</td>
</tr>
<tr>
<td><strong>Sampling-based methods</strong></td>
<td>136</td>
</tr>
<tr>
<td>Forward sampling</td>
<td>136</td>
</tr>
<tr>
<td>The accept-reject sampling method</td>
<td>137</td>
</tr>
</tbody>
</table>