<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>122</td>
</tr>
<tr>
<td>Appendix: Identifying Indiscernibles</td>
<td>122</td>
</tr>
<tr>
<td>Questions</td>
<td>123</td>
</tr>
<tr>
<td>References</td>
<td>123</td>
</tr>
<tr>
<td>VC Dimension</td>
<td>125</td>
</tr>
<tr>
<td>Approximation and Estimation Errors</td>
<td>125</td>
</tr>
<tr>
<td>Shattering</td>
<td>126</td>
</tr>
<tr>
<td>VC Dimension</td>
<td>127</td>
</tr>
<tr>
<td>Learning Result</td>
<td>128</td>
</tr>
<tr>
<td>Some Examples</td>
<td>129</td>
</tr>
<tr>
<td>Application to Neural Nets</td>
<td>132</td>
</tr>
<tr>
<td>Summary</td>
<td>133</td>
</tr>
<tr>
<td>Appendix: VC Dimension and Popper Dimension</td>
<td>133</td>
</tr>
<tr>
<td>Questions</td>
<td>134</td>
</tr>
<tr>
<td>References</td>
<td>135</td>
</tr>
<tr>
<td>Infinite VC Dimension</td>
<td>137</td>
</tr>
<tr>
<td>A Hierarchy of Classes and Modified PAC Criterion</td>
<td>138</td>
</tr>
<tr>
<td>Misfit Versus Complexity Trade-Off</td>
<td>138</td>
</tr>
<tr>
<td>Learning Results</td>
<td>139</td>
</tr>
<tr>
<td>Inductive Bias and Simplicity</td>
<td>140</td>
</tr>
<tr>
<td>Summary</td>
<td>141</td>
</tr>
<tr>
<td>Appendix: Uniform Convergence and Universal Consistency</td>
<td>141</td>
</tr>
<tr>
<td>Questions</td>
<td>142</td>
</tr>
<tr>
<td>References</td>
<td>143</td>
</tr>
<tr>
<td>The Function Estimation Problem</td>
<td>144</td>
</tr>
<tr>
<td>Estimation</td>
<td>144</td>
</tr>
<tr>
<td>Success Criterion</td>
<td>145</td>
</tr>
<tr>
<td>Best Estimator: Regression Function</td>
<td>146</td>
</tr>
<tr>
<td>Summary</td>
<td>147</td>
</tr>
<tr>
<td>Appendix: Regression Toward the Mean</td>
<td>147</td>
</tr>
<tr>
<td>Questions</td>
<td>148</td>
</tr>
<tr>
<td>References</td>
<td>149</td>
</tr>
<tr>
<td>Learning Function Estimation</td>
<td>150</td>
</tr>
<tr>
<td>Review of the Function Estimation/Regression Problem</td>
<td>150</td>
</tr>
<tr>
<td>Nearest Neighbor Rules</td>
<td>151</td>
</tr>
<tr>
<td>Kernel Methods</td>
<td>151</td>
</tr>
<tr>
<td>Neural Network Learning</td>
<td>152</td>
</tr>
<tr>
<td>Estimation with a Fixed Class of Functions</td>
<td>153</td>
</tr>
<tr>
<td>Shattering, Pseudo-Dimension, and Learning</td>
<td>154</td>
</tr>
<tr>
<td>Conclusion</td>
<td>156</td>
</tr>
<tr>
<td>Appendix: Accuracy, Precision, Bias, and Variance in Estimation</td>
<td>156</td>
</tr>
</tbody>
</table>
Questions p. 157
References p. 158
Simplicity p. 160
Simplicity in Science p. 160
Explicit Appeals to Simplicity p. 160
Is the World Simple? p. 161
Mistaken Appeals to Simplicity p. 161
Implicit Appeals to Simplicity p. 161
Ordering Hypotheses p. 162
Two Kinds of Simplicity Orderings p. 162
Two Examples p. 163
Curve Fitting p. 163
Enumerative Induction p. 164
Simplicity as Simplicity of Representation p. 165
Fix on a Particular System of Representation? p. 166
Are Fewer Parameters Simpler? p. 167
Pragmatic Theory of Simplicity p. 167
Simplicity and Global Indeterminacy p. 168
Summary p. 169
Appendix: Basic Science and Statistical Learning Theory p. 169
Questions p. 170

Support Vector Machines p. 172
Mapping the Feature Vectors p. 173
Maximizing the Margin p. 175
Optimization and Support Vectors p. 177
Implementation and Connection to Kernel Methods p. 179
Details of the Optimization Problem p. 180
Rewriting Separation Conditions p. 180
Equation for Margin p. 181
Slack Variables for Nonseparable Examples p. 181
Reformulation and Solution of Optimization p. 182
Summary p. 183
Appendix: Computation p. 184
Questions p. 185
References p. 186
Boosting p. 187
Weak Learning Rules p. 187
Combining Classifiers p. 188
Distribution on the Training Examples p. 189
The Adaboost Algorithm p. 190