

Decision Trees

Knowledge in Trees and Trees from Data

Sridhar Seshadri

Contents



Discriminative Classifiers

Rule Based Classifiers

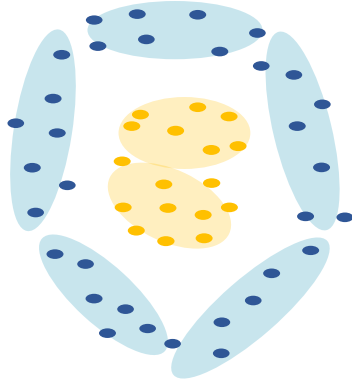
Learning Rules from Data

Rattle examples

Descriptive vs. Discriminative

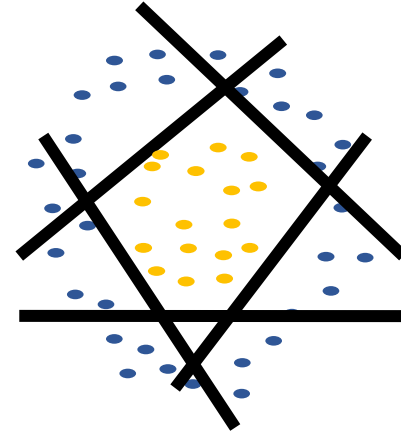


Two Types of Classifier Paradigms



Descriptive Classifiers

Bayesian Classifiers
Nearest Neighbor
Parzen Window

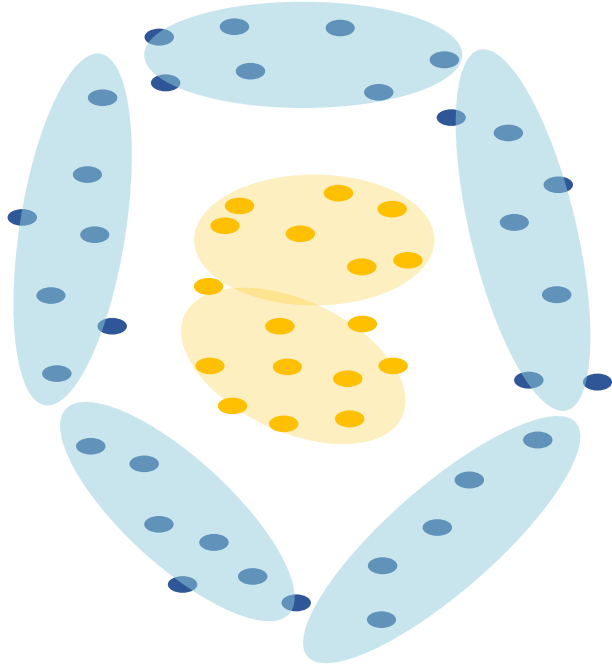


Discriminative Classifiers

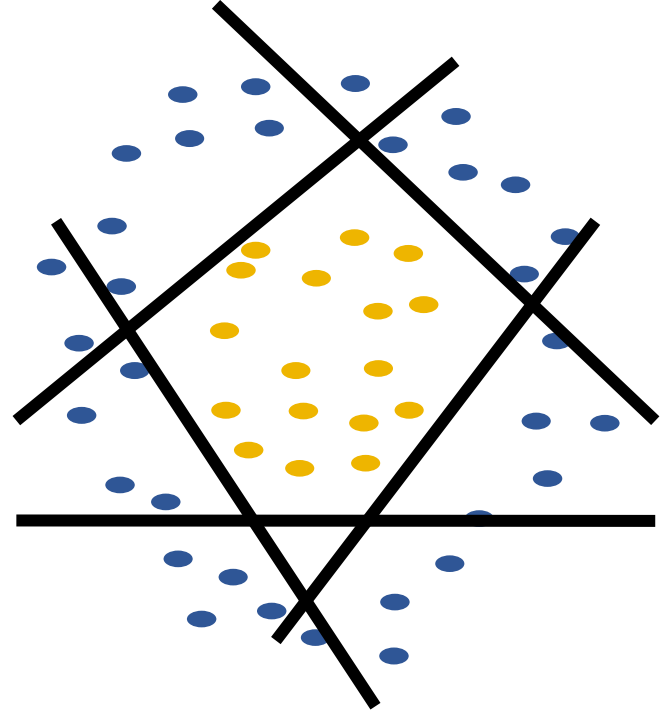
Decision Trees
Neural Networks,
Support Vector Machines

Two Types of Classifier Paradigms

I



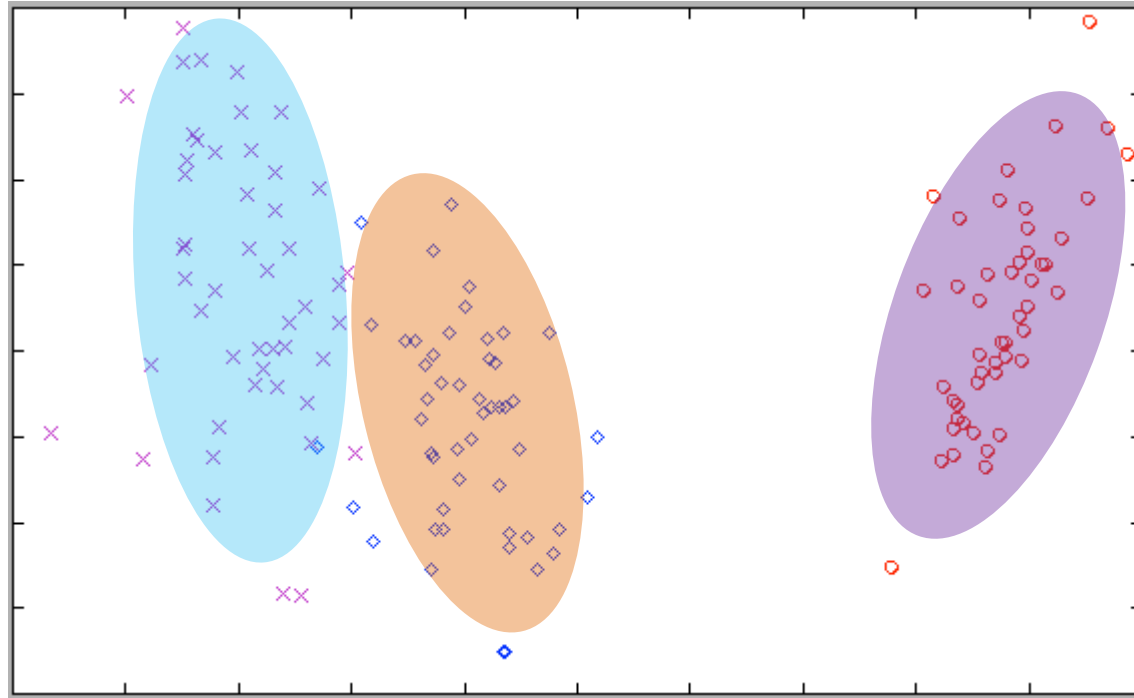
Descriptive Classifiers



Discriminative Classifiers

IRIS – Descriptive Classifier

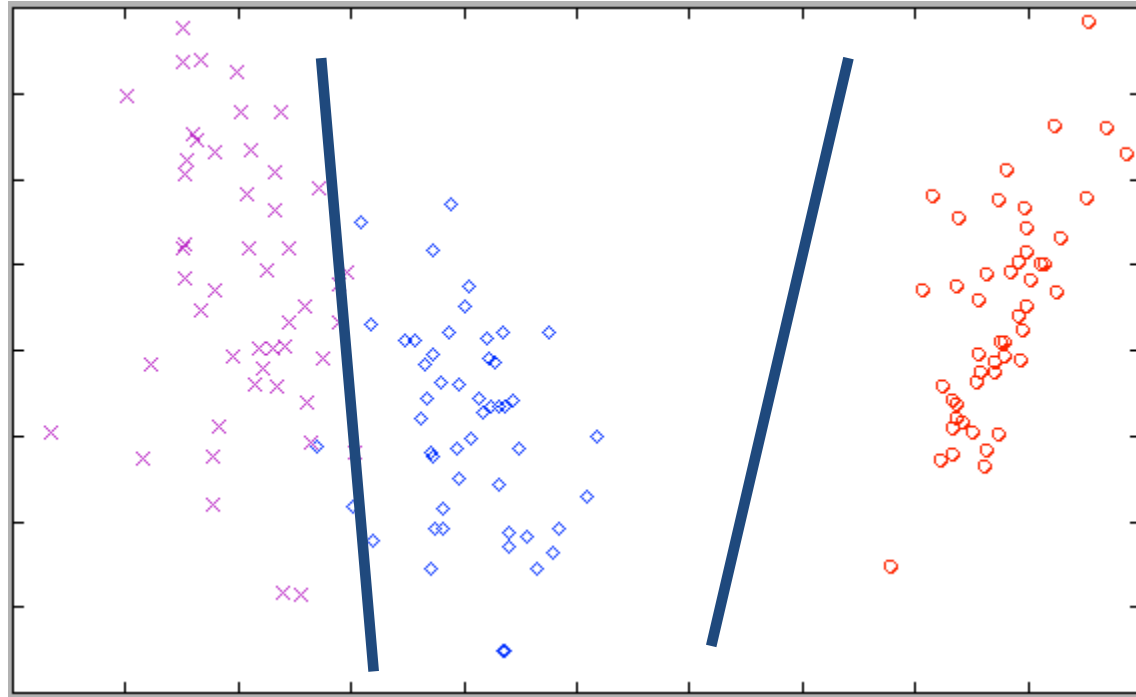
Class Density Function: $P(\mathbf{x}|c)$



IRIS – Discriminative Classifier



Class Discriminators: $\mathbf{w}^T \mathbf{x} < q$



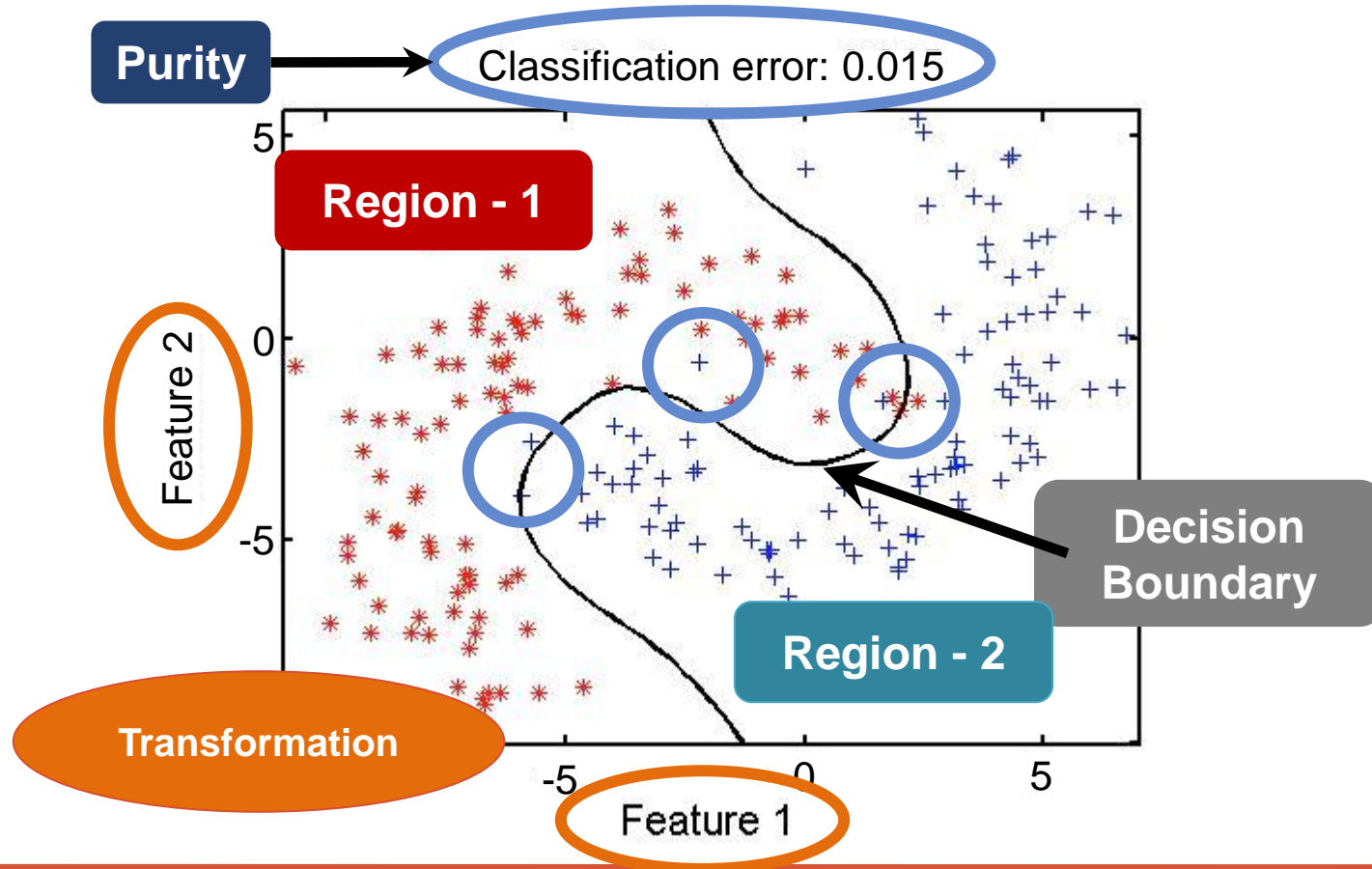
What is Discriminative “Classification”?



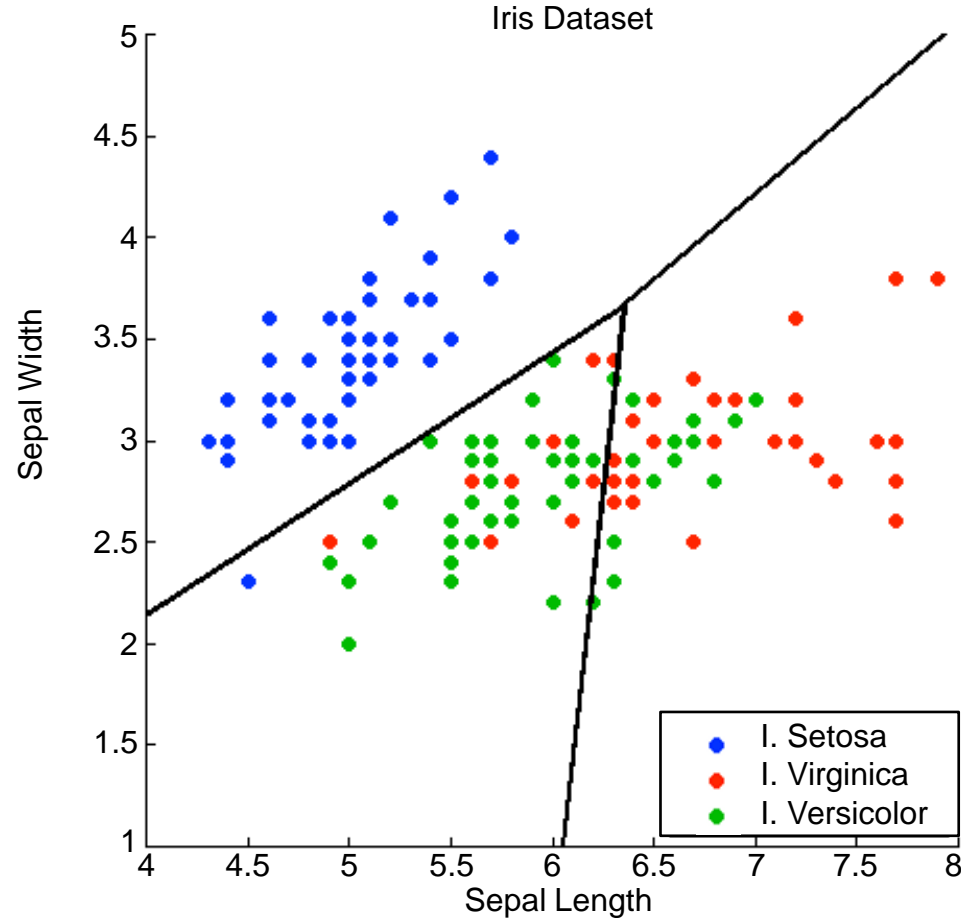
PARTITIONING the (FEATURE)
SPACE into **PURE REGIONS**
assigned to each **CLASS**

Partitioning Into Pure Region

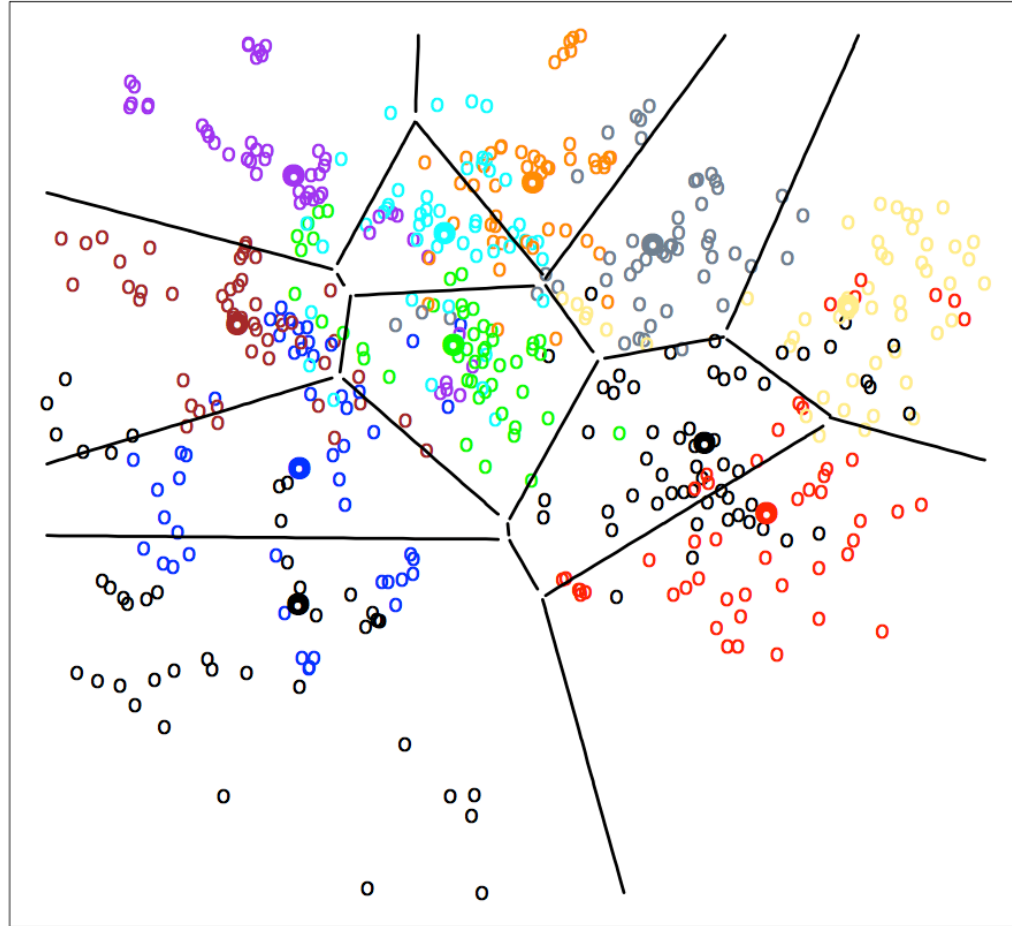
I



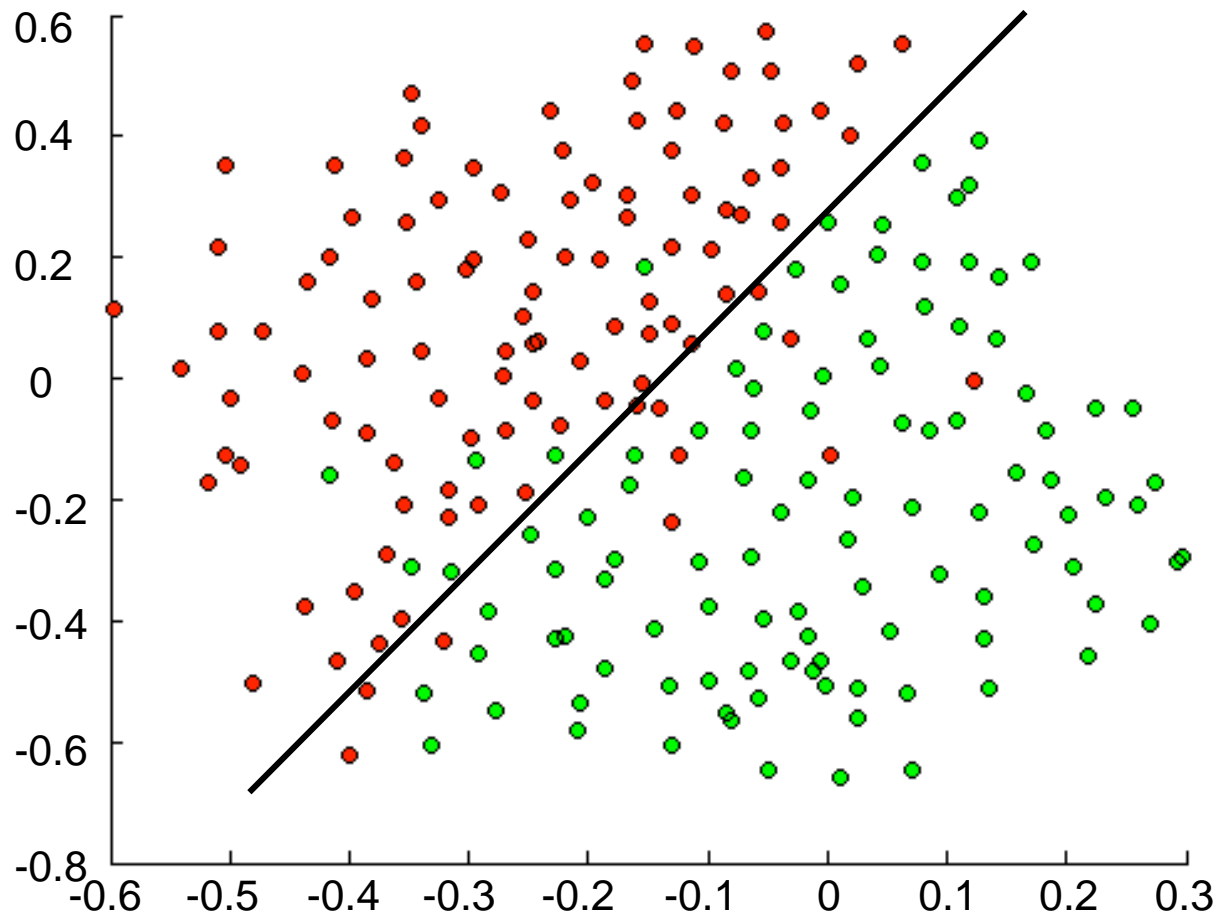
Decision Boundary for IRIS Data



Decision Boundaries for DIGITS Data

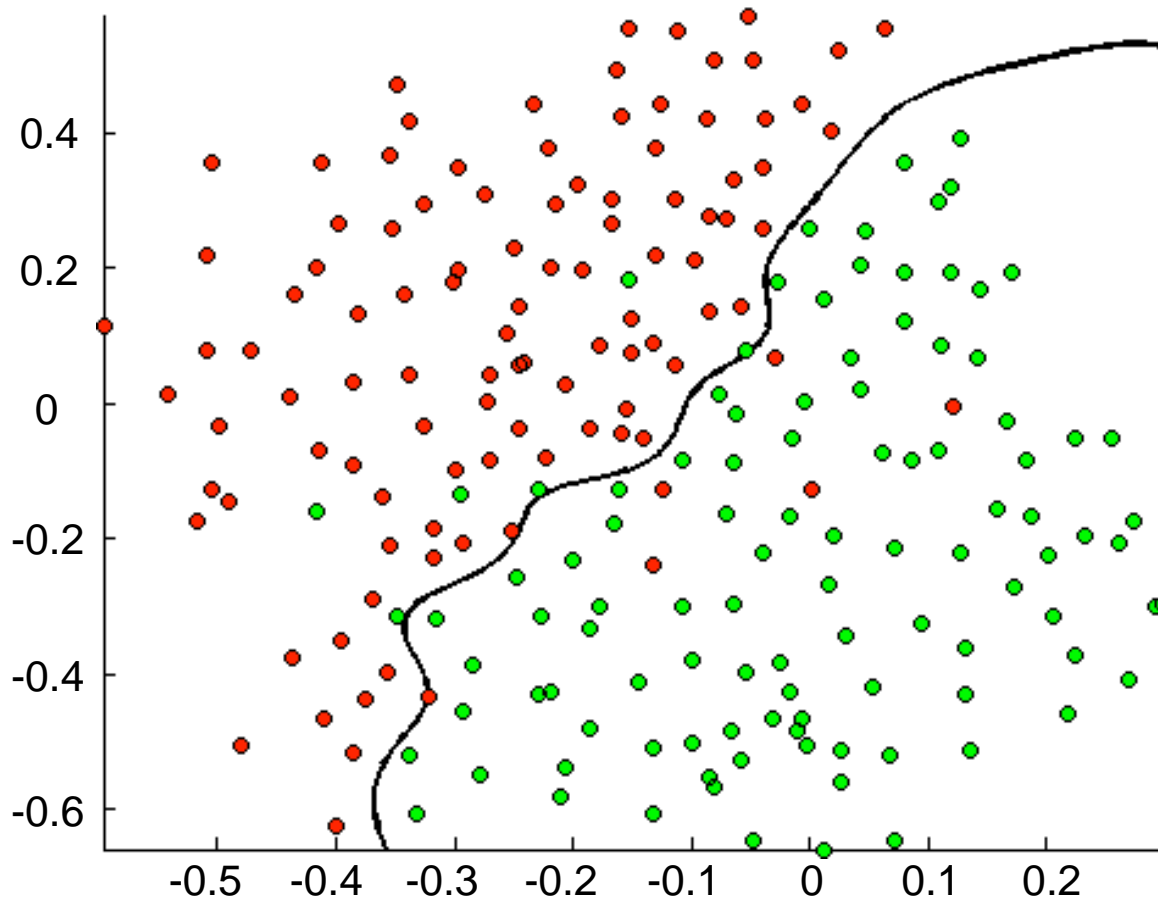


SIMPLE Decision Boundary?



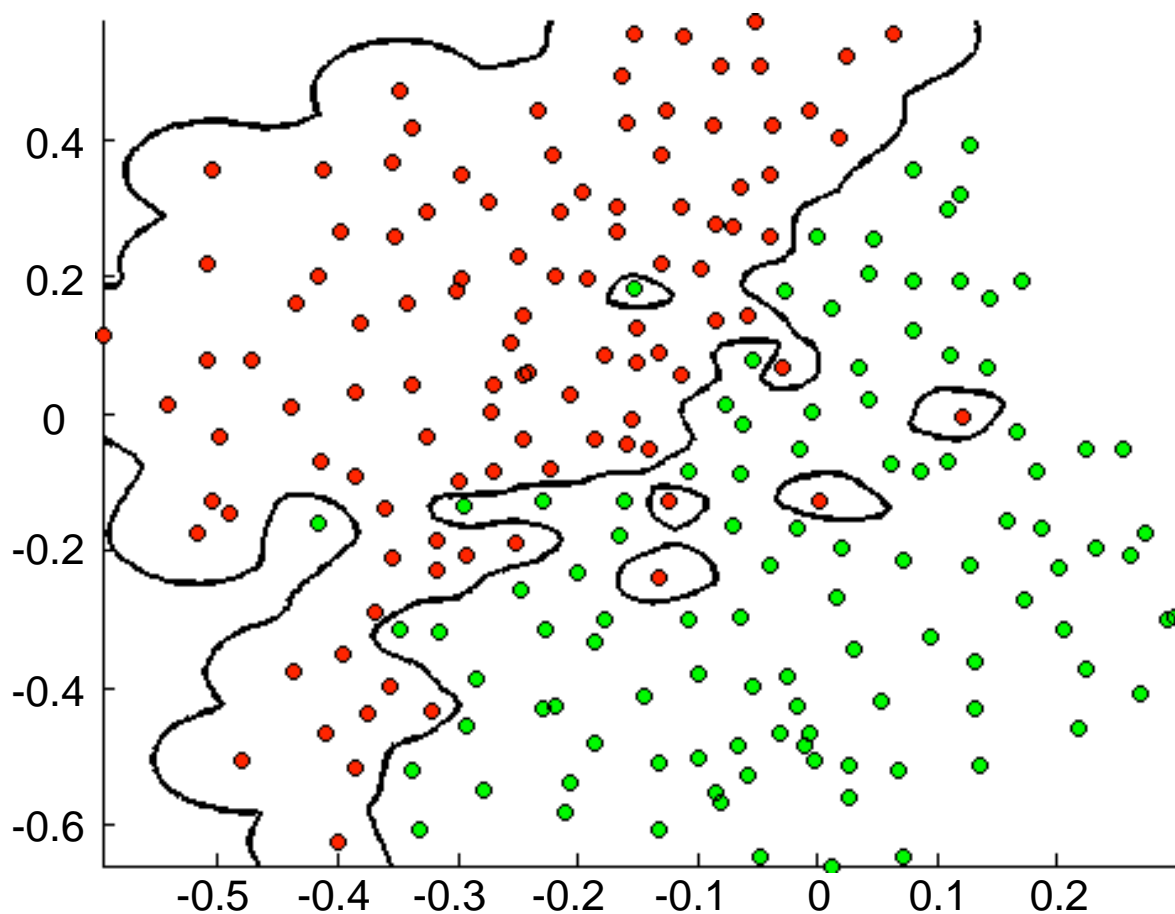
MEDIUM Decision Boundary

I

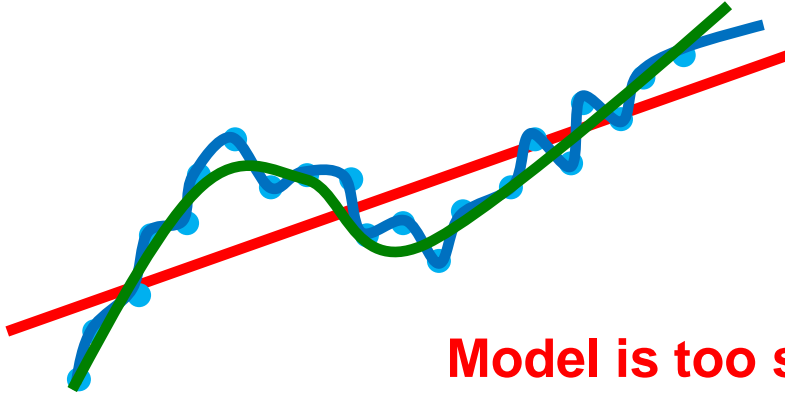


COMPLEX Decision Boundary

I



Model SIGNAL not NOISE

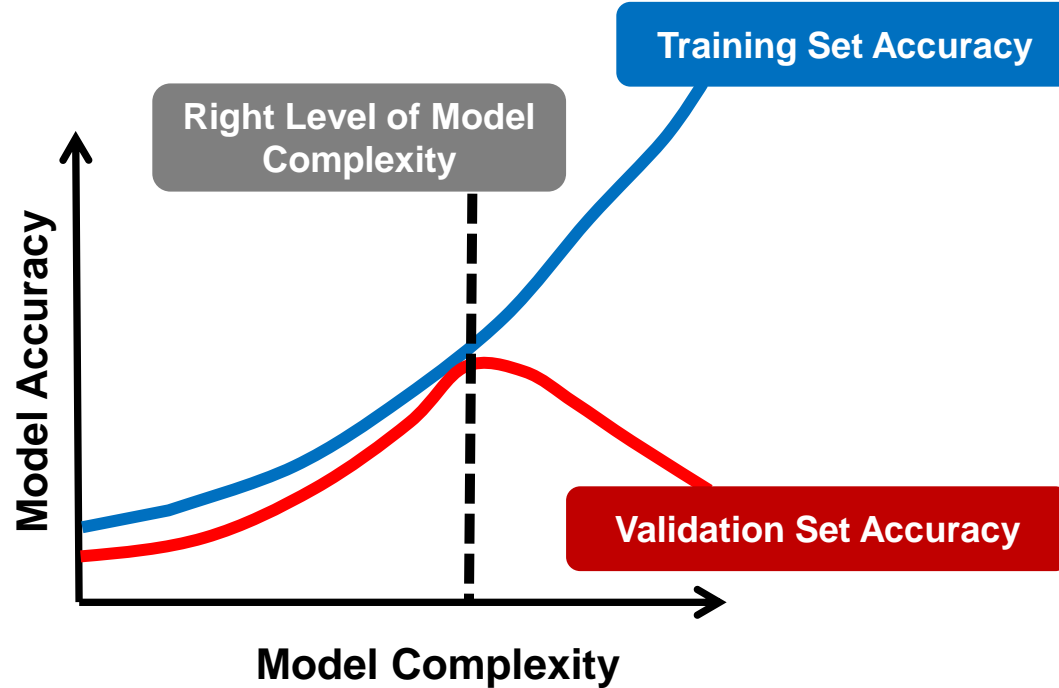


Model is too simple → UNDER LEARN

Model is too complex → MEMORIZE

Model is just right → GENERALIZE

Generalize, Don't Memorize!



Questions for Classification



What is the **NATURE** of classifier's **DECISION BOUNDARY**?

Depends on the classifier type!

What is the **COMPLEXITY** of classifier's **DECISION BOUNDARY**?

Any classifier can be made more or less complex!

How do I **CONTROL** the **COMPLEXITY** of the classifier?

What knob to use to make classifier more/less complex?

Questions for Classification



How do I know when the classifier is **COMPLEX ENOUGH**?

Degree of Noise vs. Structure in the data

Too noisy → Less complexity

How to pick the right **CLASSIFIER** to use?

Nature of the data - structured/unstructured, numeric/categorical

Complexity of the decision surface needed!

Rule Based Classifiers



Domain Knowledge Based



Fever Rules

If TEMP \leq 98.6 \rightarrow **NORMAL**

If 98.6 \leq TEMP \leq 100.0 \rightarrow **MILD-FEVER**

If 100 \leq TEMP $<$ 102.0 \rightarrow **MEDIUM-FEVER**

If TEMP $>$ 104.0 \rightarrow **HIGH-FEVER**

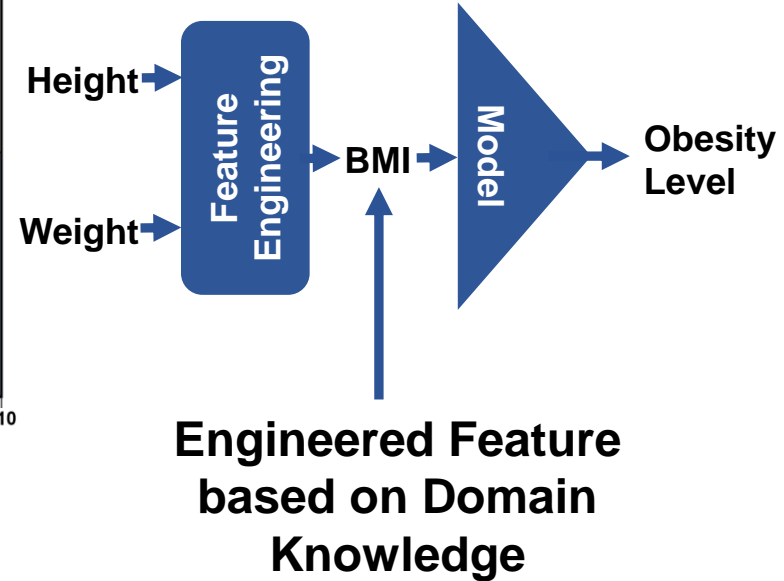
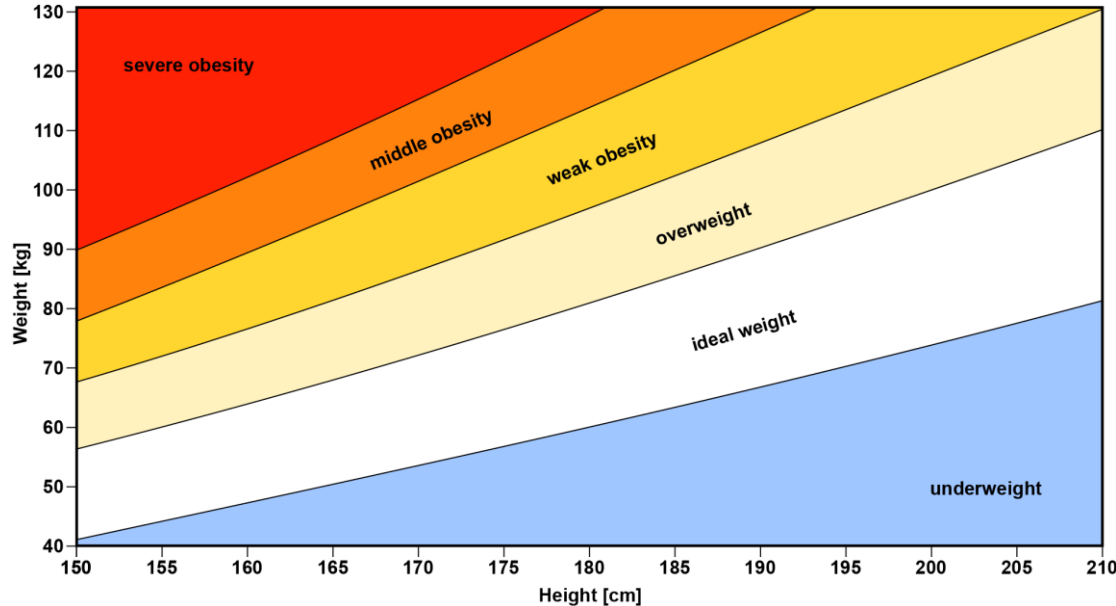
These can also be called “expert” systems.



Action Suggested
Good
Excellent

HbA _{1c}	Mean Blood Glucose	
Test Score	mg/dL	mmol/L
14.0	380	21.1
13.0	350	19.3
12.0	315	17.4
11.0	280	15.6
10.0	250	13.7
9.0	215	11.9
8.0	180	10.0
7.0	150	8.2
6.0	115	6.3
5.0	80	4.7
4.0	50	2.6

Obesity Classifier



Elements Classifier



<div>1</div> <div>H</div> <div>Hydrogen</div>																	<div>2</div> <div>He</div> <div>Helium</div>																																																																																																																				
	<div>3</div> <div>Li</div> <div>Lithium</div>	<div>4</div> <div>Be</div> <div>Beryllium</div>	<div>11</div> <div>Na</div> <div>Sodium</div>	<div>12</div> <div>Mg</div> <div>Magnesium</div>	<div>19</div> <div>K</div> <div>Potassium</div>	<div>20</div> <div>Ca</div> <div>Calcium</div>	<div>21</div> <div>Sc</div> <div>Scandium</div>	<div>22</div> <div>Ti</div> <div>Titanium</div>	<div>23</div> <div>V</div> <div>Vanadium</div>	<div>24</div> <div>Cr</div> <div>Chromium</div>	<div>25</div> <div>Mn</div> <div>Manganese</div>	<div>26</div> <div>Fe</div> <div>Iron</div>	<div>27</div> <div>Co</div> <div>Cobalt</div>	<div>28</div> <div>Ni</div> <div>Nickel</div>	<div>29</div> <div>Cu</div> <div>Copper</div>	<div>30</div> <div>Zn</div> <div>Zinc</div>		<div>31</div> <div>Ga</div> <div>Gallium</div>	<div>32</div> <div>Ge</div> <div>Germanium</div>	<div>33</div> <div>As</div> <div>Arsenic</div>	<div>34</div> <div>Se</div> <div>Selenium</div>	<div>35</div> <div>Br</div> <div>Bromine</div>	<div>36</div> <div>Kr</div> <div>Krypton</div>																																																																																																														
																								<div>37</div> <div>Rb</div> <div>Rubidium</div>	<div>38</div> <div>Sr</div> <div>Strontium</div>	<div>39</div> <div>Y</div> <div>Yttrium</div>	<div>40</div> <div>Zr</div> <div>Zirconium</div>	<div>41</div> <div>Nb</div> <div>Niobium</div>	<div>42</div> <div>Mo</div> <div>Molybdenum</div>	<div>43</div> <div>Tc</div> <div>Technetium</div>	<div>44</div> <div>Ru</div> <div>Ruthenium</div>	<div>45</div> <div>Rh</div> <div>Rhodium</div>	<div>46</div> <div>Pd</div> <div>Palladium</div>	<div>47</div> <div>Ag</div> <div>Silver</div>	<div>48</div> <div>Cd</div> <div>Cadmium</div>	<div>49</div> <div>In</div> <div>Indium</div>	<div>50</div> <div>Sn</div> <div>Tin</div>	<div>51</div> <div>Sb</div> <div>Antimony</div>	<div>52</div> <div>Te</div> <div>Tellurium</div>	<div>53</div> <div>I</div> <div>Iodine</div>	<div>54</div> <div>Xe</div> <div>Xenon</div>																																																																																												
																																										<div>55</div> <div>Cs</div> <div>Caesium</div>	<div>56</div> <div>Ba</div> <div>Barium</div>	<div>57</div> <div>La</div> <div>Lanthanum</div>	<div>72</div> <div>Hf</div> <div>Hafnium</div>	<div>73</div> <div>Ta</div> <div>Tantalum</div>	<div>74</div> <div>W</div> <div>Tungsten</div>	<div>75</div> <div>Re</div> <div>Rhenium</div>	<div>76</div> <div>Os</div> <div>Osmium</div>	<div>77</div> <div>Ir</div> <div>Iridium</div>	<div>78</div> <div>Pt</div> <div>Platinum</div>	<div>79</div> <div>Au</div> <div>Gold</div>	<div>80</div> <div>Hg</div> <div>Mercury</div>	<div>81</div> <div>Tl</div> <div>Thallium</div>	<div>82</div> <div>Pb</div> <div>Lead</div>	<div>83</div> <div>Bi</div> <div>Bismuth</div>	<div>84</div> <div>Po</div> <div>Polonium</div>	<div>85</div> <div>At</div> <div>Astatine</div>	<div>86</div> <div>Rn</div> <div>Radon</div>																																																																										
																																																												<div>87</div> <div>Fr</div> <div>Francium</div>	<div>88</div> <div>Ra</div> <div>Radium</div>	<div>89</div> <div>Ac</div> <div>Actinium</div>	<div>104</div> <div>Rf</div> <div>Rutherfordium</div>	<div>105</div> <div>Db</div> <div>Dubnium</div>	<div>106</div> <div>Sg</div> <div>Seaborgium</div>	<div>107</div> <div>Bh</div> <div>Bohrium</div>	<div>108</div> <div>Hs</div> <div>Hassium</div>	<div>109</div> <div>Mt</div> <div>Meitnerium</div>	<div>110</div> <div>Ds</div> <div>Darmstadtium</div>	<div>111</div> <div>Rg</div> <div>Roentgenium</div>	<div>112</div> <div>Uub</div> <div>Ununbium</div>	<div>113</div> <div>Uut</div> <div>Ununtrium</div>	<div>114</div> <div>Uuq</div> <div>Ununquadium</div>	<div>115</div> <div>Uup</div> <div>Ununpentium</div>	<div>116</div> <div>Uuh</div> <div>Ununhexium</div>	<div>117</div> <div>Uus</div> <div>Ununseptium</div>	<div>118</div> <div>Uuo</div> <div>Ununoctium</div>																																																								
																																																																														<div>58</div> <div>Ce</div> <div>Cerium</div>	<div>59</div> <div>Pr</div> <div>Praseodymium</div>	<div>60</div> <div>Nd</div> <div>Neodymium</div>	<div>61</div> <div>Pm</div> <div>Promethium</div>	<div>62</div> <div>Sm</div> <div>Samarium</div>	<div>63</div> <div>Eu</div> <div>Europium</div>	<div>64</div> <div>Gd</div> <div>Gadolinium</div>	<div>65</div> <div>Tb</div> <div>Terbium</div>	<div>66</div> <div>Dy</div> <div>Dysprosium</div>	<div>67</div> <div>Ho</div> <div>Holmium</div>	<div>68</div> <div>Er</div> <div>Erbium</div>	<div>69</div> <div>Tm</div> <div>Thulium</div>	<div>70</div> <div>Yb</div> <div>Ytterbium</div>	<div>71</div> <div>Lu</div> <div>Lutetium</div>																																										
																																																																																												<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																												
																																																																																																										<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>														
																																																																																																																								<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>
<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																																																								
														<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																																										
																												<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																												
																																										<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																														
																																																								<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																
																																																																						<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																		
																																																																																				<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																				
																																																																																																		<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																						
																																																																																																																<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>								
<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																																																								
														<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																																										
																												<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																																												
																																										<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																														
																																																								<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																																
																																																																						<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																																		
																																																																																				<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div>																																				
																																																																																																		<div>90</div> <div>Th</div> <div>Thorium</div>	<div>91</div> <div>Pa</div> <div>Protactinium</div>	<div>92</div> <div>U</div> <div>Uranium</div>	<div>93</div> <div>Np</div> <div>Neptunium</div>	<div>94</div> <div>Pu</div> <div>Plutonium</div>	<div>95</div> <div>Am</div> <div>Americium</div>	<div>96</div> <div>Cm</div> <div>Curium</div>	<div>97</div> <div>Bk</div> <div>Berkelium</div>	<div>98</div> <div>Cf</div> <div>Californium</div>	<div>99</div> <div>Es</div> <div>Einsteinium</div>	<div>100</div> <div>Fm</div> <div>Fermium</div>	<div>101</div> <div>Md</div> <div>Mendelevium</div>	<div>102</div> <div>No</div> <div>Nobelium</div>	<div>103</div> <div>Lr</div> <div>Lawrencium</div> </																						

Decision Trees Learning Rules From Data **I**

“Toy” Dataset

Data is in Decision_Tree_Ex.csv



Problem Description:

Ponting Flower uses five features A, B, C, D, and E to describe flowers. For example, feature A corresponds to fragrance. Each feature can take two values, for example, delicate fragrance is labeled A1 and intense fragrance as A2. The values for features B, C, D and E are {B1, B2}, {C1, C2} ..., {E1, E2}. If a particular combination is liked by at least 35% of the customers then it is labeled as green, otherwise it is labeled as red. The file contains data for 111 flowers.

A	B	C	D	E	COLOR
A1	B1	C2	D1	E2	GREEN
A2	B2	C2	D1	E1	RED
A2	B2	C1	D1	E1	GREEN
A1	B2	C1	D1	E2	GREEN
A1	B1	C2	D2	E2	GREEN
A2	B2	C2	D1	E1	RED

Process of Building a Decision Tree From Data



Partitioning and recursive partitioning

Purity measures

Which Feature to Pick First?



Idea is to compute the purity gain from partitioning

Entropy of a region can be calculated as: $\sum -p(c) * \log_2(p(c))$ where the summation is taken over all classes and $p(c)$: Probability of a class in that region

The definition of purity is $1 - \text{entropy}$ (if binary classification)

An Algorithm



Step 1: Calculate the entropy for the overall dataset

Step 2: Calculate the entropy for splitting using each attribute.

Step 3: Calculate the Purity Gain for splitting using each attribute using the previous formula.

An Algorithm



Step 4: Select the attribute with the highest Purity Gain and split the data on the basis of this attribute.

Step 5: Repeat Step 1 to Step 4 until all the data is classified

Sample Steps



Step 1: Calculate the entropy for the overall dataset:

Total Number of Observation: $n(\text{Total}) = 110$

Number of Observations with Green class: $n(\text{Green}) = 89$

Probability for Green Class: $p(\text{Green}) = 89/110 = 0.8091$

$-p(\text{Green}) * \text{Binary Logarithm}(p(\text{Green})) = -0.8091 * -0.306 = 0.2473$

Number of Observations with Red class: $n(\text{Red}) = 21$

Probability for Red Class: $p(\text{Red}) = 21/110 = 0.1909$

$-p(\text{Red}) * \text{Binary Logarithm}(p(\text{Red})) = -0.1909 * -2.3890 = 0.4561$

Entropy (Overall Data) $= 0.2473 + 0.4561 = 0.7034$ (i.e., Entropy (0))

Step 2



Calculate Entropy for splitting using each attribute. Take for example, attribute A

We have two categorical values A1 and A2. So to calculate Entropy for attribute A [entropy(A,COLOR)] have to calculate Entropy of A1 and A2 on the basis of Color and then multiply it by the fraction of samples that have feature A1 and A2, labeled as $p(A1)$ and $p(A2)$.

Entropy (A, COLOR) =

$$p(A1) * \text{Entropy} (A1, \text{COLOR}) + p(A2) * \text{Entropy} (A2, \text{COLOR})$$

$$\text{Entropy (A, COLOR)} = 0.4545 * 0 + 0.5455 * 0.934068 = 0.509492$$

$$\text{Purity Gain for Attribute A} = \text{Entropy (COLOR)} - \text{Entropy (A, COLOR)}$$

$$\text{Purity Gain (A, COLOR)} = 0.7034 - 0.509492 = 0.193878$$

Step 3:

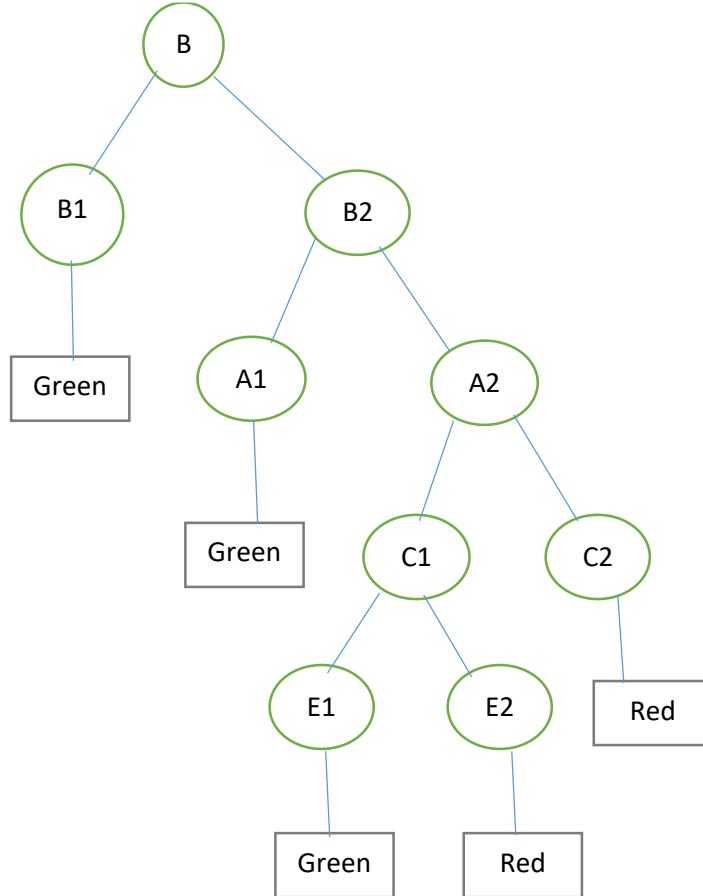


Repeat for all attributes

	Entropy	Purity Gain
Overall	0.703369	
A	0.509492	0.193878
B	0.453166	0.250203
C	0.695062	0.008307
D	0.695654	0.007715
E	0.693644	0.009725

Split on B

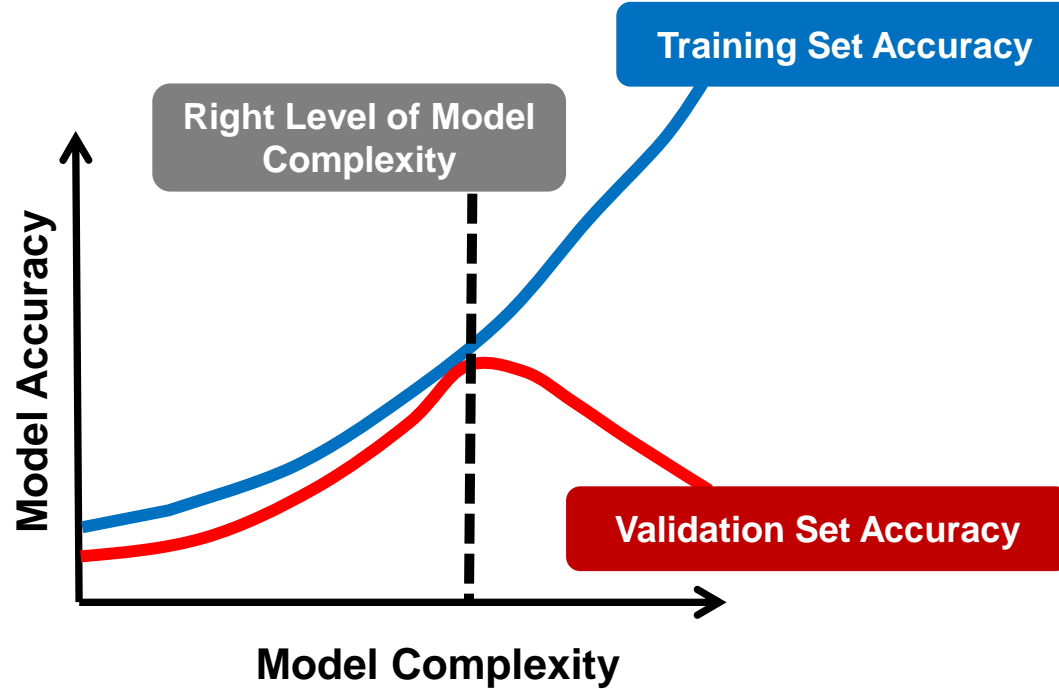
Final Tree



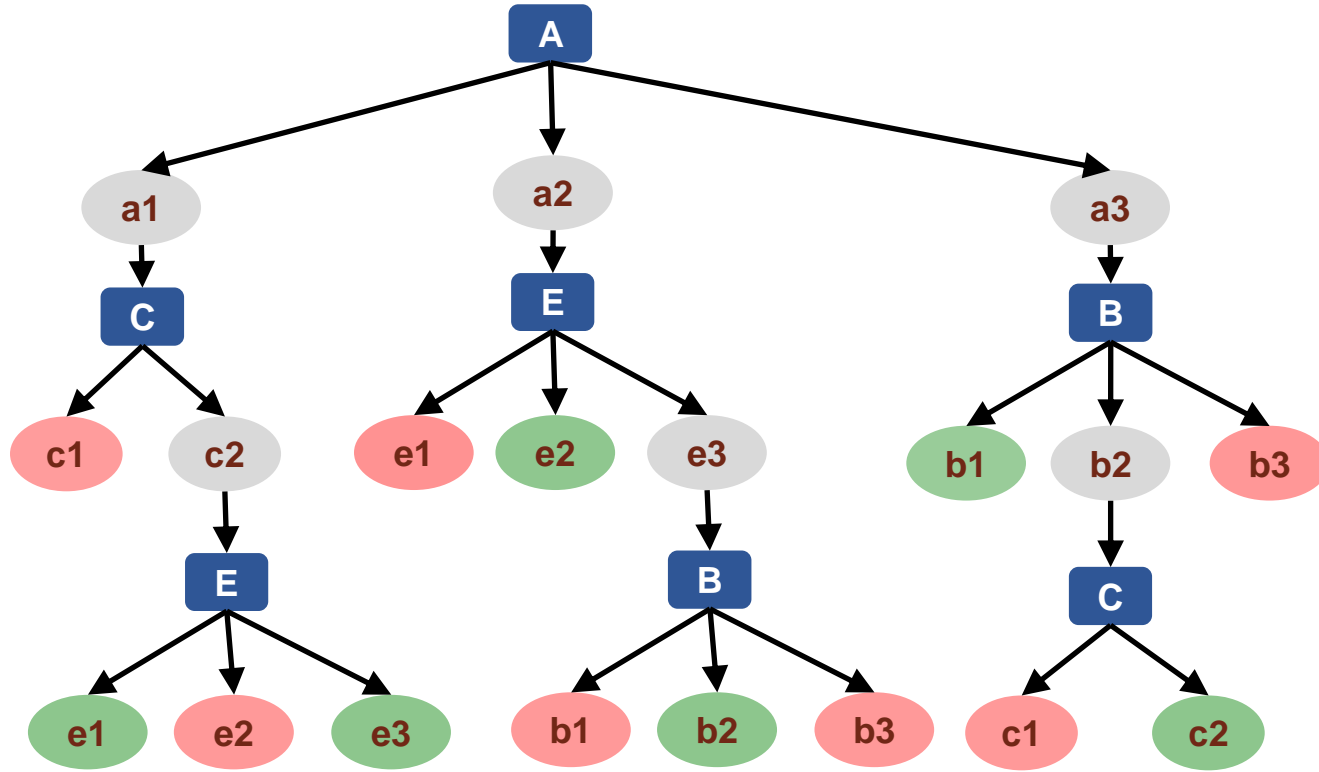
What are the rules?

There are 5!

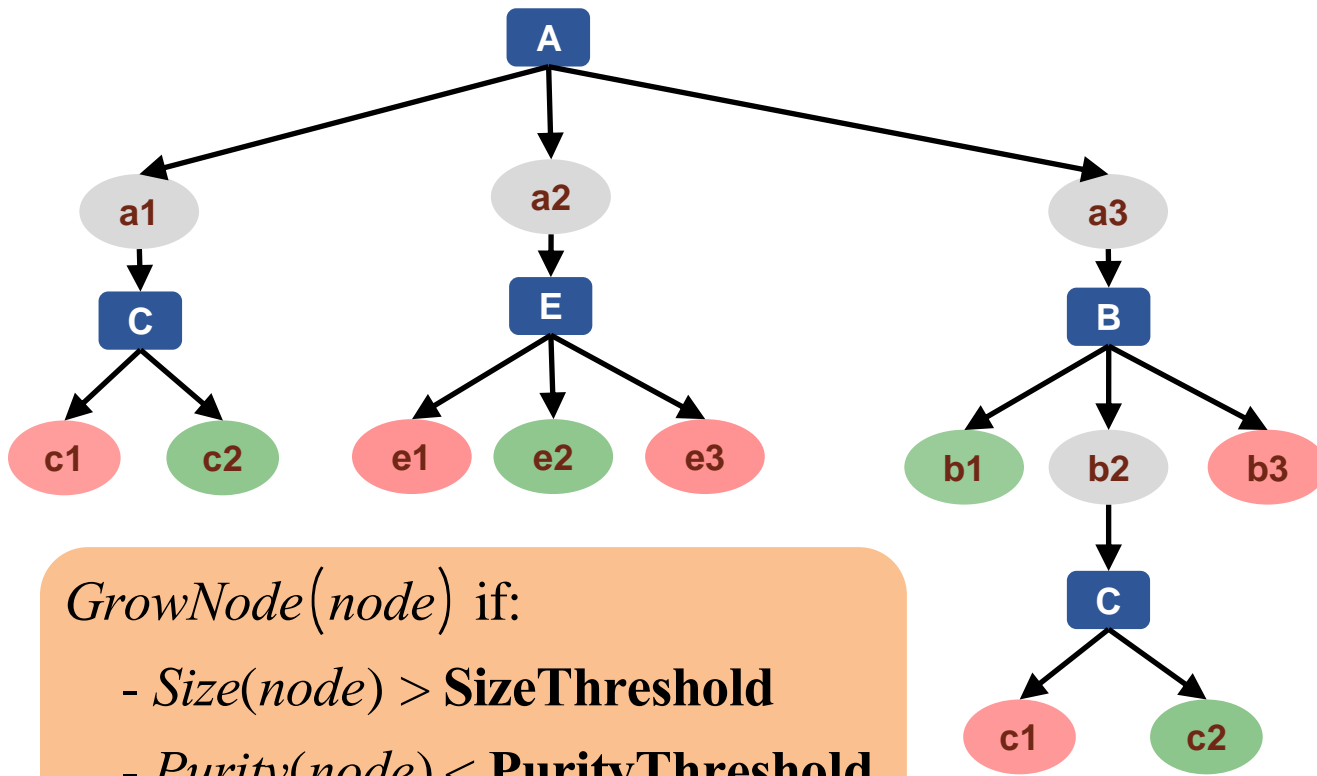
Controlling COMPLEXITY



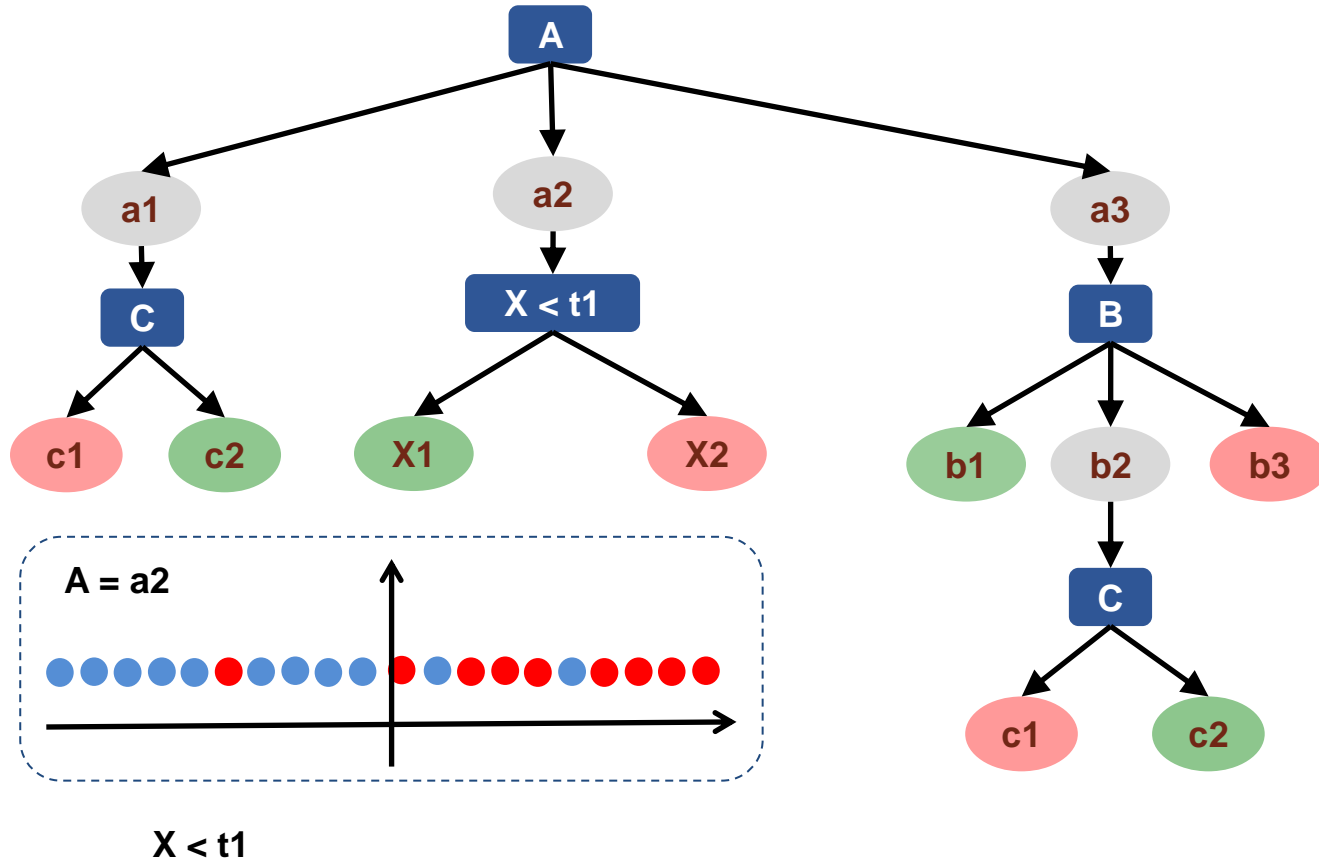
Example of Overfitting



Pruning



Numeric Attributes



Decision Tree vs Clustering (Kmeans/Hclust)



Decision trees are also used for clustering; observations in the same leaf make a cluster.

The difference between decision tree and clustering techniques is that decision trees can be applied on labelled data only.

Thus, whenever target values (labels) are available, we should prefer decision trees to clustering methods(unsupervised).

Type of Decision Tree



Classification Tree – when target value is a class

Regression Tree – When target value is numeric

We will discuss examples of both in Rattle

Classification Tree – Toy Example



Objective – To classify flowers as liked by at least 35% vs not liked (Green vs. Red).

R Data Miner - [Rattle (Decision_Tree_Ex.csv)]

Project Tools Settings Help

Rattle Version 5.2.0 togaware

Execute New Open Save Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

Source: ☒ File ☐ ARFF ☐ ODBC ☐ R Dataset ☐ RData File ☐ Library ☐ Corpus ☐ Script

Filename: Decision_Tree_Ex... Separator: , Decimal: . ☒ Header

☒ Partition 70/15/15 Seed: 42 View Edit

☒ Input ☐ Ignore Weight Calculator:

Target Data Type
☒ Auto ☐ Categorical ☐ Numeric ☐ Survival

No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	A	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
2	B	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
3	C	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
4	D	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
5	E	Categorical	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2
6	COLOR	Categorical	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 2

Running Decision Tree

Output slightly different in R version 3.6.3



Execute New Open Save Export Stop Quit

Data Explore Test Transform Cluster Associate **Model** Evaluate Log

Type: ☒ Tree ☐ Forest ☐ Boost ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ All

Target: COLOR Algorithm: ☒ Traditional ☐ Conditional

Min Split: 5 Max Depth: 4 Priors:

Min Bucket: 2 Complexity: 0.0000 Loss Matrix:

```
1) root 77 11 GREEN (0.8571429 0.1428571)
 2) B=B1 44 0 GREEN (1.0000000 0.0000000) *
 3) B=B2 33 11 GREEN (0.6666667 0.3333333)
 6) A=A1 17 0 GREEN (1.0000000 0.0000000) *
 7) A=A2 16 5 RED (0.3125000 0.6875000)
 14) E=E1 8 3 GREEN (0.6250000 0.3750000)
 28) C=C1 5 0 GREEN (1.0000000 0.0000000) *
 29) C=C2 3 0 RED (0.0000000 1.0000000) *
 15) E=E2 8 0 RED (0.0000000 1.0000000) *
```

Classification tree:
rpart(formula = COLOR ~ ., data = crs\$dataset[crs\$train, c(crs\$input, crs\$target)], method = "class", model = TRUE, parms = list(split = "information"), control = rpart.control(minsplit = 5, minbucket = 2, maxdepth = 4, cp = 0.00001, usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] A B C E

Root node error: 11/77 = 0.14286

n= 77

	CP	nsplit	rel error	xerror	xstd
1	0.27273	0	1.00000	1.00000	0.27915
2	0.22727	2	0.45455	0.90909	0.26816
3	0.00001	4	0.00000	0.18182	0.12688

Time taken: 0.00 secs

Rattle timestamp: 2020-04-19 16:08:37 ashis

Complexity Table

Output slightly different in R version 3.6.3



Root node error: $11/77 = 0.14286$

$n = 77$

	CP	Nsplit	Rel error	Xerror	Xstd
1	0.27273	0	1.0000	1.0000	0.27915
2	0.22727	2	0.4545	0.90909	0.26816
3	0.00001	4	0.0000	0.18182	0.12688

Classification Rules – Post Model-fitting

Data Explore Test Transform Cluster Associate Model Evaluate Log

Type: ☒ Tree ☐ Forest ☐ Boost ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ All

Target: COLOR Algorithm: ☒ Traditional ☐ Conditional Model Builder: rpart

Min Split: 5 Max Depth: 4 Priors: ☐ Include Missing

Min Bucket: 2 Complexity: 0.0000 Loss Matrix: **Rules** Draw

Rattle timestamp: 2020-03-16 20:09:35 ashis

Tree as rules:

```

Rule number: 15 [COLOR=RED cover=8 (10%) prob=1.00]
  B=B2
  A=A2
  E=E2

Rule number: 29 [COLOR=RED cover=3 (4%) prob=1.00]
  B=B2
  A=A2
  E=E1
  C=C2

Rule number: 28 [COLOR=GREEN cover=5 (6%) prob=0.00]
  B=B2
  A=A2
  E=E1
  C=C1

Rule number: 6 [COLOR=GREEN cover=17 (22%) prob=0.00]
  B=B2
  A=A1

Rule number: 2 [COLOR=GREEN cover=44 (57%) prob=0.00]
  B=B1

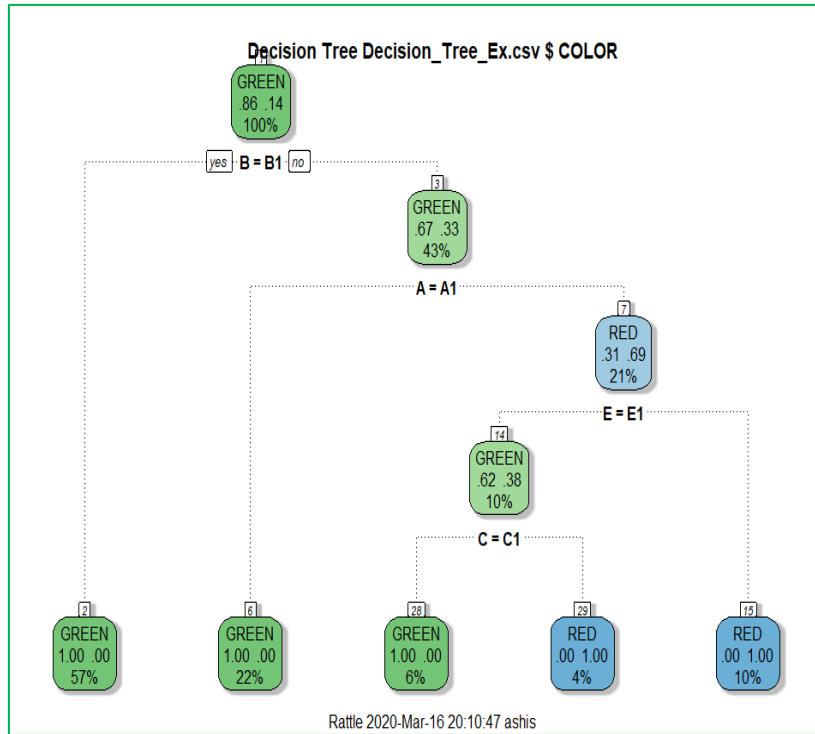
[1] 9 8 5 6 3 1 7 4 2
  
```

Rattle timestamp: 2020-03-16 20:09:36 ashis

Classification Tree – Tree

Output slightly different in R version 3.6.3

I



Tree as rules:

```
Rule number: 15 [COLOR=RED cover=8 (10%) prob=1.00]
B=B2
A=A2
E=E2

Rule number: 29 [COLOR=RED cover=3 (4%) prob=1.00]
B=B2
A=A2
E=E1
C=C2

Rule number: 28 [COLOR=GREEN cover=5 (6%) prob=0.00]
B=B2
A=A2
E=E1
C=C1

Rule number: 6 [COLOR=GREEN cover=17 (22%) prob=0.00]
B=B2
A=A1

Rule number: 2 [COLOR=GREEN cover=44 (57%) prob=0.00]
B=B1
```

[1] 9 8 5 6 3 1 7 4 2

Rattle timestamp: 2020-03-16 20:09:36 ashis

Tuning Parameters



Method

Minsplit – minimum observations that must exist at a node before it is split

Minbucketsize – minimum number of observations at a leaf node

Max_Depth – maximum depth of the tree

CP - controls the complexity of the tree

Classification Tree – Stop at Depth = 2



GUI Screenshot showing the 'Evaluate' tab selected. The 'Model' is set to 'Tree' and 'Data' is set to 'Validation'. The output displays the error matrix for the Decision Tree model on Decision_Tree_Ex.csv [validate] (co).

Model: ☒ Tree ☐ Boost ☐ Forest ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐

Data: ☐ Training ☒ Validation ☐ Testing ☐ Full ☐ Enter ☐ CSV File

Risk Variable: **Report:** ☒

Error matrix for the Decision Tree model on Decision_Tree_Ex.csv [validate] (co)

	Predicted		
Actual	GREEN	RED	Error
GREEN	11	0	0
RED	0	5	0

Error matrix for the Decision Tree model on Decision_Tree_Ex.csv [validate] (pr)

	Predicted		
Actual	GREEN	RED	Error
GREEN	68.8	0.0	0
RED	0.0	31.2	0

Overall error: 0%, Averaged class error: 0%

Rattle timestamp: 2020-04-11 11:19:31 sridhar

What If the Target Variable is Numeric Instead of a Class Category ?



Whenever, the target variable is numeric, we use Regression Tree in place of Classification Tree.

We discuss an example of Regression Tree for predicting the price of used cars on the Toyota dataset.

Regression Tree – Toyota Car



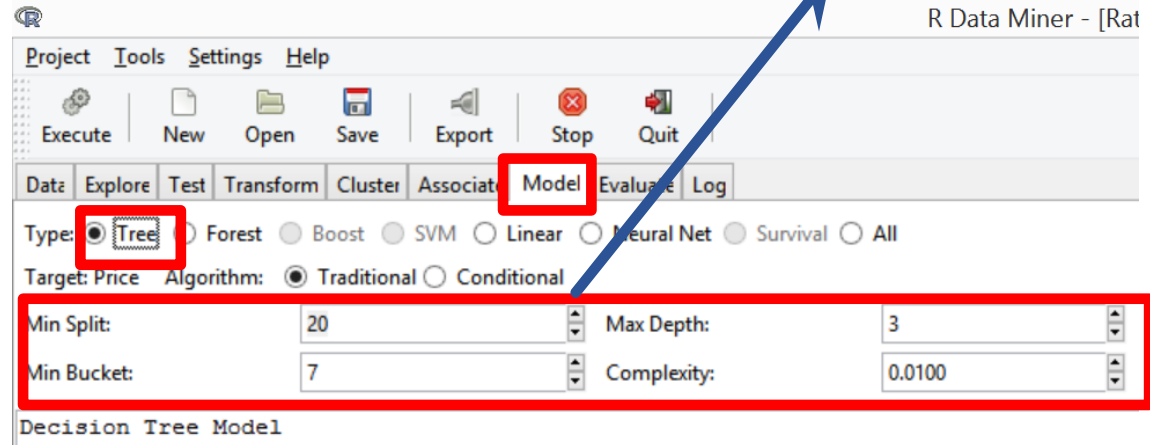
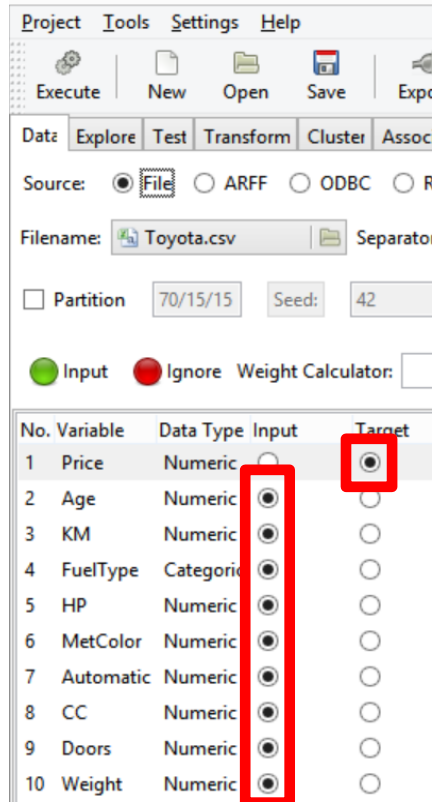
Objective – Predict the price of used car based on the its characteristics

Variable	Description
Price	Offer price in euros
Age	Age in months as of August 2004
Kilometers	Accumulated kilometers on odometer
Fuel type	Fuel type (<i>Petrol, Diesel, CNG</i>)
Horse Power	Horsepower
Metallic	Metallic color? (Yes = 1, No = 0)
Automatic	Automatic (Yes = 1, No = 0)
CC	Cylinder volume in cubic centimeters
Doors	Number of doors
Weight	Weight in kilograms

Top five rows

Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight
13500	23	46986	Diesel	90	1	0	2000	3	1165
13750	23	72937	Diesel	90	1	0	2000	3	1165
13950	24	41711	Diesel	90	1	0	2000	3	1165
14950	26	48000	Diesel	90	0	0	2000	3	1165
13750	30	38500	Diesel	90	0	0	2000	3	1170

Regression Tree to Predict the “Price” of Used Car



Source: Rattle GUI / Togaware;

Download data from <https://www.biz.uiowa.edu/faculty/jledolter/datamining/datatext.html>

Regression Tree – Rules (Post Model-Fitting)

R Data Miner - [Rattle (Toyota.csv)]

Project Tools Settings Help

Execute New Open Save Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

Type: ☒ Tree ☐ Forest ☐ Boost ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ All

Target: Price Algorithm: ☒ Traditional ☐ Conditional

Min Split: 20 Max Depth: 3 Priors:

Min Bucket: 7 Complexity: 0.0100 Loss Matrix:

Model Build

☐ Include Missing **Rules** Draw

Rattle timestamp: 2020-04-11 11:35:23 sridhar

=====

Tree as rules:

Rule number: 9 [Price=9338.40136054422 cover=441 (31%)]
 Age>=32.5
 Age>=56.5
 Age< 68.5

Rule number: 8 [Price=7925.56887755102 cover=392 (27%)]
 Age>=32.5
 Age>=56.5
 Age>=68.5

Rule number: 10 [Price=10728.44 cover=225 (16%)]
 Age>=32.5
 Age< 56.5
 Age>=44.5

Rule number: 11 [Price=12274.9739583333 cover=192 (13%)]
 Age>=32.5
 Age< 56.5
 Age< 44.5

Rule number: 12 [Price=16842.98 cover=100 (7%)]
 Age< 32.5
 Weight< 1278
 Weight< 1122

Regression Tree

Output different in R version 3.6.3

I

Execute New Open Save Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

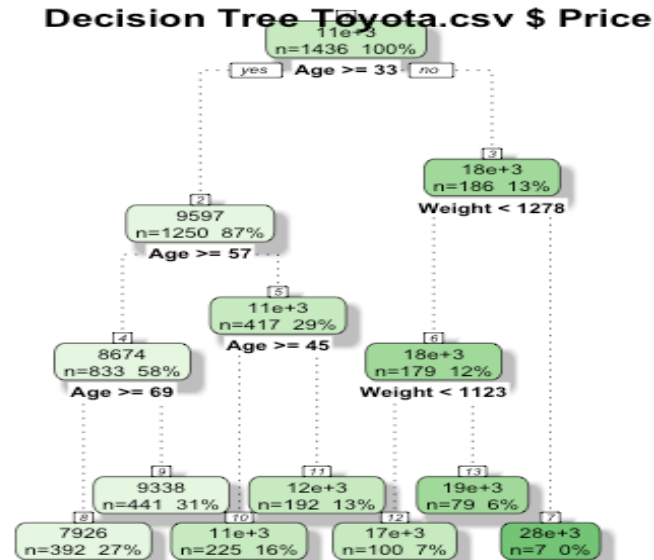
Type: ☒ Tree ☐ Forest ☐ Boost ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ All

Target: Price Algorithm: ☒ Traditional ☐ Conditional

Min Split: 20 Max Depth: 3 Priors:

Min Bucket: 7 Complexity: 0.0100 Loss Matrix:

Model Build: ☐ Include Missing



Rattle 2020-Apr-11 11:38:07 sridhar

Regression Tree – Evaluation (Post Model-fitting): Predicted Price

I

Type: ☐ Error Matrix ☐ Risk ☐ Cost Curve ☐ Hand ☐ Lift ☐ ROC ☐ Precision ☐ Sensitivity ☐ Pr v Ob ☒ Score

Model: ☒ Tree ☐ Boost ☐ Forest ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ KMeans ☐ HClust

Data: ☒ Training ☐ Validation ☐ Testing ☐ Full ☐ Enter ☐ CSV File ☐ R Dataset

Risk Variable: Report: ☒ Class ☐ Probability Include: ☐ Identifiers ☒ All

Score Files

Name: Toyota_train_score_all.csv

Save in folder: SS

We kept 100% data for training

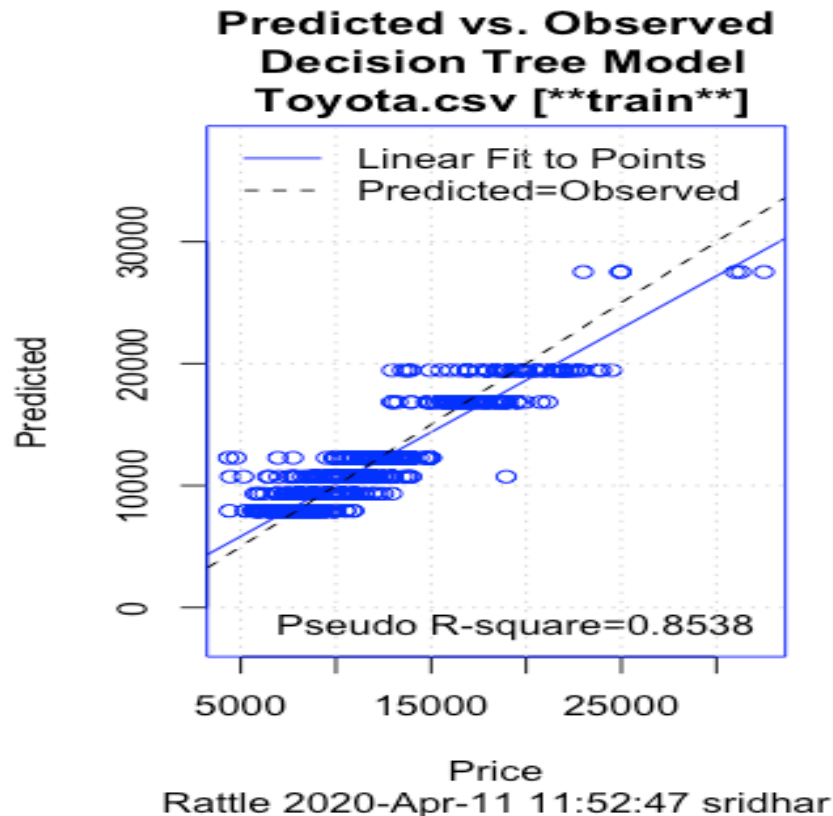
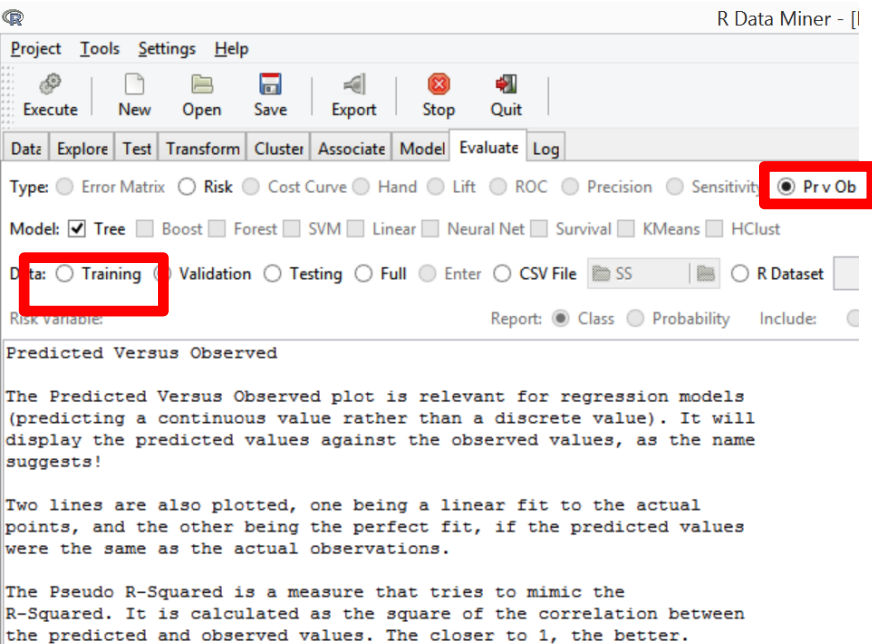
Predicted price

Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight	rpart
22750	30	34000	Petrol	192	1	0	1800	3	1185	19453
17950	24	21716	Petrol	110	1	0	1600	3	1105	16843
16750	24	25563	Petrol	110	0	0	1600	3	1065	16843
16950	30	64359	Petrol	110	1	0	1600	3	1105	16843
15950	30	67660	Petrol	110	1	0	1600	3	1105	16843
13690	29.4	75648.5	Petrol	77.2	0.8	0	1520	3	1065	15799

Predicted vs Observed

Output slightly different in R version 3.6.3

I



From Trees to Forest



Randomly select data

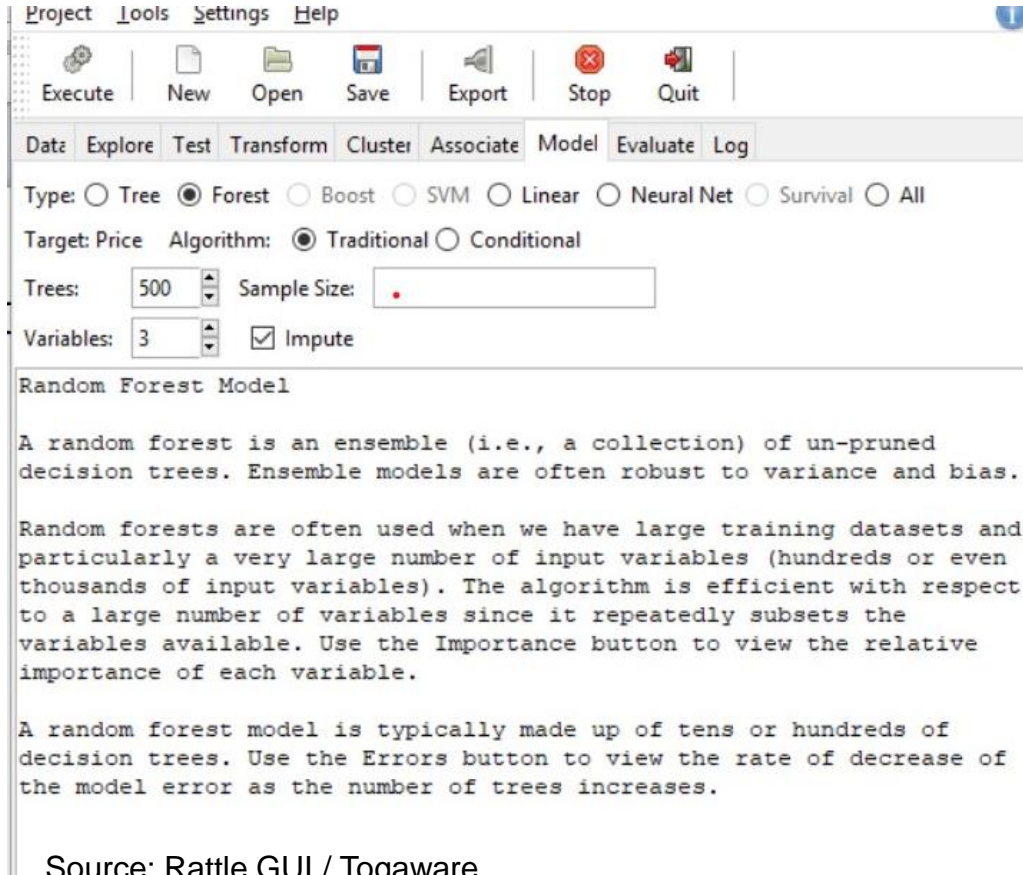
Randomly select variables on which to split

Grow many trees

Use a “ensemble” approach to predict

Robust and improves predictability

Random Forest Application Example



Very simple options:
Number of Trees
Variables to select at split

Compare With the Decision Tree

Output slightly different in latest version.



GUI interface for the Random Forest Model (Rattle GUI / Togaware).

Type: ☐ Tree ☒ Forest ☐ Boost ☐ SVM ☐ Linear ☐ Neural Net ☐ Survival ☐ All

Target: Price Algorithm: ☒ Traditional ☐ Conditional Model Builder:

Trees: 500 Sample Size: Importance Rules 1

Variables: 3 ☒ Impute Errors OOB ROC

Summary of the Random Forest Model

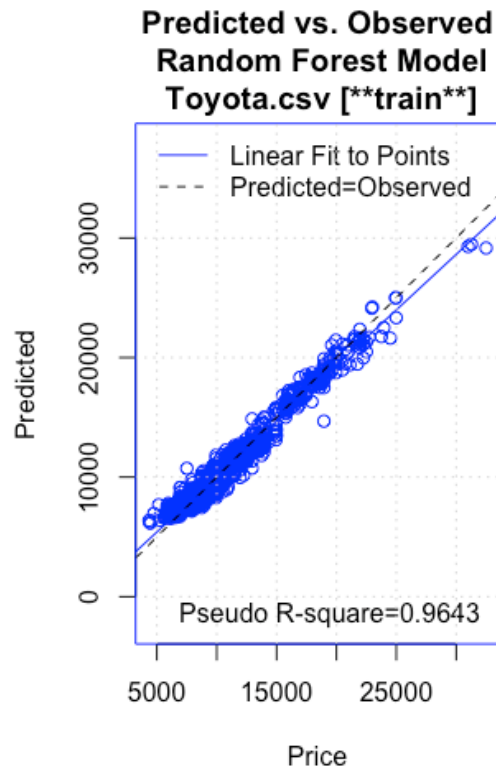
Number of observations used to build the model: 1436
Missing value imputation is active.

Call:
randomForest(formula = Price ~ .,
data = crs\$dataset[, c(crs\$input, crs\$target)],
ntree = 500, mtry = 3, importance = TRUE, replace = FALSE, na.action = randomForest::na.roughfix)

Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 3
Mean of squared residuals: 1184371
% Var explained: 90.99

Variable Importance

	%IncMSE	IncNodePurity
Age	69.67	6031088243



Source: Rattle GUI / Togaware

Rattle 2020-Apr-11 11:57:58 sridhar

Importance of Variables

```

variables: 3 ☐ impute
Summary of the Random Forest Model
=====
Number of observations used to build the model:
Missing value imputation is active.

Call:
  randomForest(formula = Price ~ .,
               data = crs$dataset[, c(crs$input,
                                     ntree = 500, mtry = 3, importance =
                                     Type of random forest: regression
                                     Number of trees: 500
No. of variables tried at each split: 3

               Mean of squared residuals: 1184371
               % Var explained: 90.99

Variable Importance
=====
      %IncMSE IncNodePurity
Age          69.67      6031088243
Weight       29.87      1622266817
KM           26.00      2373253314
HP           21.08      689671295
CC           20.36      386400930
Doors        14.06      93706341
FuelType      7.26      110407155
Automatic     6.74      18151149
MetColor      4.07      49019327

Time taken: 1.44 secs

Rattle timestamp: 2020-04-11 11:56:10 sridhar
=====

```


Summary



Decision trees are intuitive

Overfitting and variable selection issues

Greedy method and does not question whether the improvement is statistically significant

Summary



Predictive power might be poor

Conditional tree

Random Forest

Further Readings



Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, & C.J. (1984). *Classification and Regression Trees*. Taylor and Francis.

Buhlmann, P. (2010) [Remembrance of Leo Breiman](#). *The Annals of Applied Statistics*, 4(4).

Exercise – Part 1

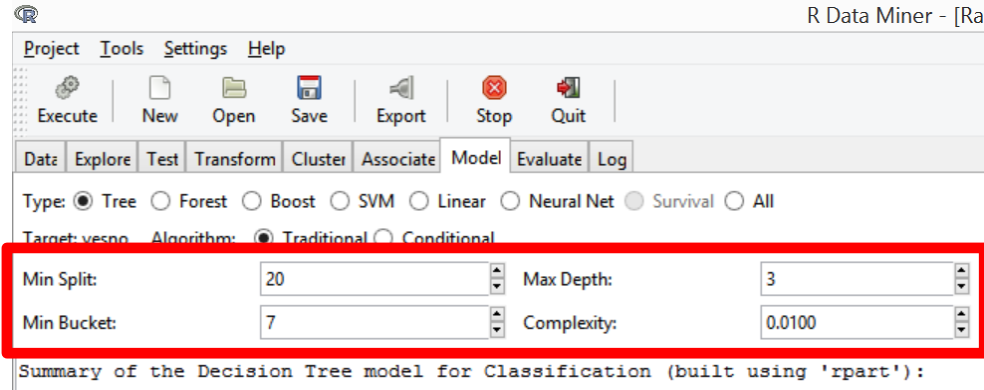
Source- DAAG package library R

Original source - <http://archive.ics.uci.edu/ml/machine-learning-databases/spambase/>



Use Classification Tree to make a Spam Filter

1) Make a spam filter on the data from **spam.csv** file. Here are variable descriptions



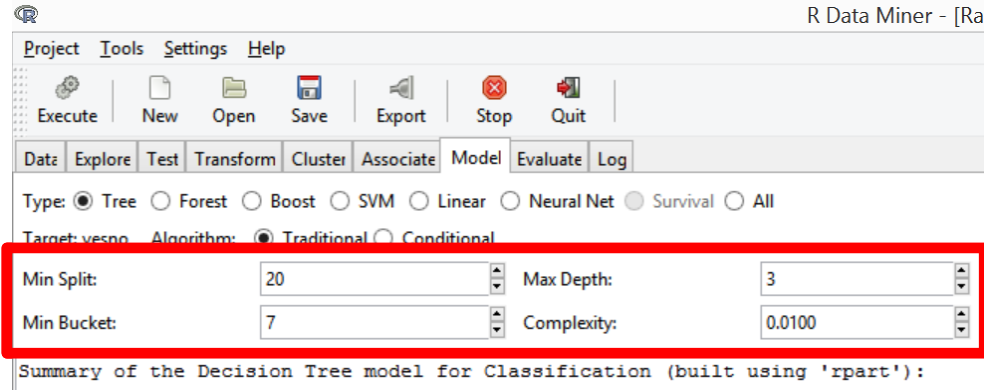
cr1.tot total length of words in capitals
dollar number of occurrences of the \\$ symbol
bang number of occurrences of the ! symbol
money number of occurrences of the word 'money'
n000 number of occurrences of the string '000'
make number of occurrences of the word 'make'
yesno outcome variable, a factor with levels n not spam, y spam

Hint – Run a model with the default values in Rattle and note the training and validation error rates. Then change one of Min Split, Max Depth, Min Bucket, Complexity at a time, keeping other three constants and notice the changes in the training and validation error rates.

Exercise – Part 2



2) Briefly discuss how each of these four values (min split, max. depth, min bucket, and complexity) affect bias and variance . For example, what does increasing min split value do to the bias and variance of the trained model?



crl.tot total length of words in capitals
dollar number of occurrences of the \\$ symbol
bang number of occurrences of the ! symbol
money number of occurrences of the word 'money'
n000 number of occurrences of the string '000'
make number of occurrences of the word 'make'
yesno outcome variable, a factor with levels n not spam, y spam

Hint – Run a model with the default values in Rattle and note the training and validation error rates. Then change one of these four at a time, keeping other three constants and notice the changes in the training and validation error rates.

References



“[BMI Graph](#)” by [Bibi Saint-Pol](#) is licensed under CC BY 3.0

Ledolter, J. (n.d.) [Data Mining: ToyotaCorolla.csv](#). Retrieved May 22, 2019 from University of Iowa.

Rattle GUI / Togaware
(<https://rattle.togaware.com/>)