CSE 40171: Artificial Intelligence

Probabilistic Read-Out Layers for Artificial Neural Networks: Combining Bayesian Models with Artificial Neural Networks
Homework #8 is due on 12/11 at 11:59PM
Final Project Deliverable are Due
12/18 at 11:59PM

(See Course Website for Instructions)
Quiz #2 will take place on 12/11 in class. See review checklist on course website.
Course Instructor Feedback (CIF)
Deadline: 11:59PM, 12/15/19
How do we deploy Bayes’ theorem for decision making?
Features

Feature Extractor Until This Point

Simonyan and Zisserman 2015
Decision Making

Classifier at This Point

\[ P(y = j | x) = \frac{e^{v_j(x)}}{\sum_{i=1}^{N} e^{v_i(x)}} \]

Sum over all of the classes
Decision Making Machines

Niu and Suen Pattern Recognition 2012
Even with these read-out options, why are we lacking good decision making?

Softmax, SVM, and other decision machines cannot express structure in their outputs

Most read-outs are not probabilistic (or weakly calibrated to produce probability-like outputs)

Under-appreciated problem: much of a network’s performance is in the classifier, not the feature extraction layers
Largest Constraint: Generalization

“These are goats”
Few-shot Learning

Tenenbaum et al. Science 2011
Learning Inductive Biases

Step 1.
- Cups are cup-shaped.
- Balls are round.

Step 2.
- Cups are cup-shaped.
- Balls are round.

Step 3.
- Balls are round.
- Cups are cup-shaped.
- "_____’s" are ______-shaped.

Step 4.
- Shaped thing is "giraffe".
- Shaped thing is "crayon".
- Shaped thing is "tractor".

Unsupervised Hierarchical Categorization

Tenenbaum et al. Science 2011
Salakhutdinov et al. Workshop on Unsupervised and Transfer Learning 2012

Introduced Hierarchical Nonparametric Bayesian Models for handcrafted features

Campero et al. CogSci 2017

Extends the work of Salakhutdinov et al. for use with features from deep learning
Learning a Class-Specific Similarity Metric from One Example
Complex feature spaces with a hierarchical semantic structure

Deep Boltzmann Machine

\[ M \text{ replicated softmax units with tied weights} \]

\[ \text{Multinomial unit sampled } M \text{ times} \]

Motivation: Generative Model

Salakhutdinov et al. 2013
Modern Deep Network Features

AlexNet: Krizhevsky et al. 2012

VGG-16: Simonyan and Zisserman 2015
Generative Semantic Organization

**Step 1:** Extract features from a chosen DNN

**Step 2:** Hierarchical Bayesian Model’s parameters are inferred by approximating the posterior via Markov Chain Monte Carlo methods
Generative Semantic Organization

Assume two-level hierarchy where $N$ observed inputs are partitioned into $C$ basic-level categories.

These categories are in turn partitioned into $K$ supercategories.

The distributions over features of each basic-level category are assumed to be multivariate Gaussian with a category specific mean $M_c$ and with precision terms $\tau_d^c$ that are assumed to be independent.
Conjugate Normal-Gamma prior over \( \{\mu_c, \tau_c\} \); determined by the supercategory specific level-2 parameters \( \mu_k, \tau_k, \alpha_k \).

\( \mu_k \) and \( \tau_k \) constitute the expected values of the lower-level parameters and \( \alpha_k \) controls the variability of \( \tau_c \) around its mean. For the conjugate priors over the level-2 parameters, assume Normal, Exponential and Inverse-Gamma distributions that are further shaped by parameters \( \alpha_0 \) and \( \gamma_0 \).
Bayesian Inference

Given a set of observations, the model iteratively performs Bayesian inference by alternating between sampling the parameters and inferring category assignments.

When learning distributions at each step of the iteration, supercategory membership is fixed and the parameters are sampled from posteriors that are analytically computed using the conjugate priors.

Supercategory membership for each category is learned by fixing the current parameters and the rest of the hierarchical structure. Assignment to any of the existing supercategories or to a newly created one.
MSR Cambridge Dataset

Baseline: “texture-of-textures” (handcrafted) features

<table>
<thead>
<tr>
<th>Model</th>
<th>Category: Cow</th>
<th></th>
<th>Category: Flower</th>
<th></th>
<th>Average</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 ex</td>
<td>2 ex</td>
<td>4 ex</td>
<td>20 ex</td>
<td>1 ex</td>
<td>2 ex</td>
</tr>
<tr>
<td>HB</td>
<td>0.77</td>
<td>0.81</td>
<td>0.84</td>
<td>0.89</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td>HB-Flat</td>
<td>0.62</td>
<td>0.69</td>
<td>0.80</td>
<td>0.89</td>
<td>0.59</td>
<td>0.64</td>
</tr>
<tr>
<td>HB-Var</td>
<td>0.61</td>
<td>0.73</td>
<td>0.83</td>
<td>0.89</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>Euclidean</td>
<td>0.59</td>
<td>0.61</td>
<td>0.63</td>
<td>0.66</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.83</td>
<td>0.84</td>
<td>0.87</td>
<td>0.89</td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td>MLE</td>
<td>0.58</td>
<td>0.64</td>
<td>0.78</td>
<td>0.89</td>
<td>0.55</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Salakhutdinov et al. Workshop on Unsupervised and Transfer Learning 2012
AUROC on the MSR dataset in the one-shot learning task

<table>
<thead>
<tr>
<th># Examples from Withheld Class</th>
<th>Alexnet</th>
<th>VGG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1ex</td>
<td>2ex</td>
</tr>
<tr>
<td>Oracle</td>
<td>.99</td>
<td>1</td>
</tr>
<tr>
<td>HB-Full</td>
<td>.91</td>
<td>.96</td>
</tr>
<tr>
<td>One Supercategory</td>
<td>.87</td>
<td>.94</td>
</tr>
<tr>
<td>NearestN</td>
<td>.84</td>
<td>.86</td>
</tr>
<tr>
<td>T of T*</td>
<td>.76</td>
<td>.80</td>
</tr>
</tbody>
</table>

Campero et al. CogSci 2017
AUROC on the MSR dataset with limited training data

<table>
<thead>
<tr>
<th># Training Examples</th>
<th># Examples from Withheld Class</th>
<th>1 ex</th>
<th>2 ex</th>
<th>4 ex</th>
<th>20 ex</th>
<th>1 ex</th>
<th>2 ex</th>
<th>4 ex</th>
<th>20 ex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ex</td>
<td>Alexnet</td>
<td>.87</td>
<td>.87</td>
<td>.88</td>
<td>.89</td>
<td>.90</td>
<td>.90</td>
<td>.90</td>
<td>.92</td>
</tr>
<tr>
<td>4 ex</td>
<td></td>
<td>.92</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td>.93</td>
<td>.97</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>10 ex</td>
<td></td>
<td>.92</td>
<td>.96</td>
<td>.99</td>
<td>.99</td>
<td>.92</td>
<td>.96</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>18 ex</td>
<td></td>
<td>.92</td>
<td>.95</td>
<td>.98</td>
<td>.99</td>
<td>.91</td>
<td>.96</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>All examples</td>
<td></td>
<td>.91</td>
<td>.96</td>
<td>.98</td>
<td>.99</td>
<td>.92</td>
<td>.97</td>
<td>.98</td>
<td>.99</td>
</tr>
</tbody>
</table>

Campero et al. CogSci 2017
MSR semantic tree discovered by the Full Model

Camero et al. CogSci 2017
Ground-truth Gazoobian Framework
Model’s Inferred Semantic Hierarchy of Gazoobian Objects

Campero et al. CogSci 2017