# SELF-TAUGHT LEARNING: TRANSFER LEARNING FROM UNLABELED DATA

Richard Wen
rwen@ryerson.ca

April 7, 2017

*Article Summary for Knowledge Discovery Course*
*Instructed by Dr. Cherie Ding*

Ryerson University

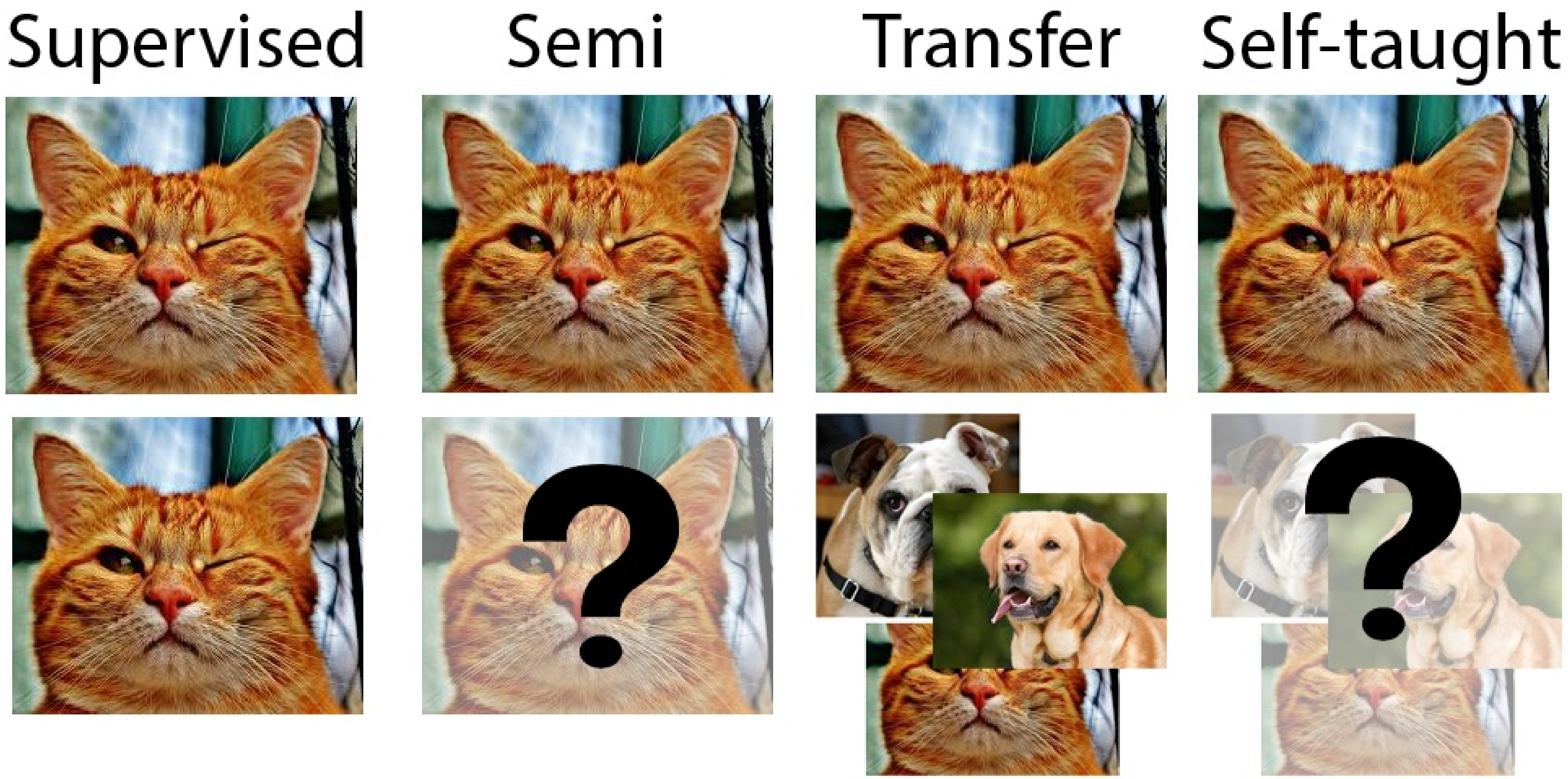
# OUTLINE

1. Introduction
2. Methods
3. Results
4. Discussion and Conclusion

Ryerson University

# INTRODUCTION

# LEARNING




Ryerson University

# WHAT WE SEE

# WHAT THE COMPUTER SEES

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 8 | 41 | 46 | 171 | 230 | 151 | 181 | 67 | 174 | 51 | 206 | 22 | 20 | 49 | 10 | 128 | 202 | 164 | 23 | 178 | 159 | 102 | 122 | 121 | 245 | 53 | 169 |
| 84 | 100 | 203 | 3 | 180 | 41 | 123 | 107 | 235 | 148 | 48 | 219 | 15 | 221 | 244 | 71 | 123 | 150 | 20 | 71 | 130 | 62 | 118 | 7 | 207 | 127 | 92 | 44 |
| 216 | 181 | 156 | 122 | 238 | 74 | 106 | 181 | 73 | 5 | 185 | 153 | 189 | 68 | 14 | 139 | 234 | 242 | 229 | 214 | 108 | 199 | 117 | 62 | 191 | 40 | 229 | 128 |
| 245 | 240 | 11 | 118 | 35 | 213 | 52 | 120 | 47 | 124 | 208 | 70 | 61 | 83 | 23 | 10 | 190 | 37 | 7 | 99 | 122 | 248 | 228 | 191 | 217 | 255 | 244 | 5 |
| 92 | 141 | 24 | 107 | 75 | 239 | 186 | 187 | 76 | 253 | 236 | 135 | 232 | 180 | 232 | 171 | 217 | 106 | 47 | 161 | 100 | 130 | 101 | 33 | 209 | 39 | 116 | 3 |
| 250 | 30 | 123 | 158 | 111 | 142 | 126 | 234 | 98 | 68 | 129 | 99 | 54 | 157 | 28 | 190 | 81 | 8 | 51 | 112 | 58 | 4 | 95 | 66 | 168 | 83 | 146 | 67 |
| 27 | 99 | 131 | 159 | 96 | 57 | 81 | 8 | 113 | 24 | 125 | 95 | 248 | 130 | 87 | 93 | 9 | 253 | 149 | 50 | 103 | 235 | 166 | 256 | 90 | 56 | 215 | 55 |
| 255 | 85 | 157 | 21 | 171 | 60 | 105 | 33 | 129 | 112 | 183 | 25 | 253 | 200 | 30 | 141 | 228 | 109 | 6 | 251 | 32 | 11 | 135 | 79 | 158 | 201 | 87 | 193 |
| 68 | 88 | 11 | 253 | 127 | 181 | 130 | 167 | 205 | 182 | 52 | 75 | 95 | 193 | 70 | 88 | 86 | 131 | 101 | 97 | 30 | 117 | 126 | 64 | 207 | 206 | 129 | 202 |
| 239 | 42 | 106 | 204 | 59 | 104 | 2 | 251 | 77 | 73 | 97 | 79 | 127 | 60 | 118 | 173 | 194 | 221 | 194 | 249 | 22 | 185 | 249 | 17 | 23 | 234 | 27 | 7 |
| 250 | 223 | 199 | 138 | 250 | 64 | 151 | 226 | 242 | 55 | 139 | 142 | 180 | 133 | 155 | 241 | 225 | 150 | 154 | 144 | 188 | 25 | 166 | 238 | 42 | 1 | 69 | 44 |
| 80 | 97 | 132 | 180 | 1 | 57 | 252 | 215 | 200 | 164 | 81 | 83 | 153 | 128 | 148 | 163 | 12 | 240 | 139 | 51 | 30 | 164 | 256 | 14 | 79 | 159 | 53 | 175 |
| 37 | 193 | 116 | 138 | 25 | 66 | 169 | 104 | 21 | 239 | 86 | 246 | 110 | 190 | 147 | 125 | 76 | 146 | 241 | 113 | 134 | 245 | 45 | 185 | 173 | 2 | 238 | 98 |
| 237 | 76 | 160 | 63 | 50 | 142 | 57 | 11 | 112 | 43 | 253 | 162 | 184 | 101 | 107 | 55 | 162 | 19 | 41 | 46 | 10 | 222 | 69 | 112 | 234 | 32 | 174 | 36 |
| 192 | 58 | 156 | 114 | 111 | 18 | 15 | 183 | 125 | 254 | 68 | 192 | 142 | 98 | 248 | 26 | 200 | 101 | 60 | 244 | 100 | 83 | 79 | 220 | 81 | 192 | 57 | 204 |
| 231 | 248 | 187 | 138 | 1 | 210 | 197 | 63 | 251 | 79 | 107 | 143 | 97 | 53 | 190 | 251 | 100 | 63 | 159 | 48 | 238 | 171 | 67 | 33 | 34 | 25 | 234 | 2 |
| 232 | 232 | 206 | 116 | 114 | 177 | 224 | 24 | 237 | 204 | 73 | 166 | 56 | 104 | 133 | 39 | 222 | 35 | 236 | 37 | 200 | 194 | 125 | 1 | 143 | 150 | 102 | 193 |
| 158 | 118 | 22 | 122 | 256 | 33 | 75 | 10 | 3 | 186 | 56 | 220 | 110 | 218 | 29 | 73 | 161 | 54 | 53 | 143 | 30 | 224 | 81 | 51 | 57 | 253 | 156 | 54 |
| 229 | 56 | 8 | 69 | 42 | 220 | 209 | 62 | 92 | 131 | 95 | 140 | 179 | 11 | 130 | 19 | 37 | 180 | 16 | 6 | 181 | 99 | 135 | 205 | 61 | 224 | 73 | 181 |
| 124 | 250 | 8 | 192 | 196 | 25 | 221 | 233 | 24 | 123 | 89 | 190 | 18 | 180 | 85 | 4 | 195 | 95 | 103 | 48 | 90 | 212 | 102 | 33 | 151 | 228 | 113 | 248 |
| 95 | 186 | 224 | 49 | 214 | 76 | 220 | 153 | 135 | 247 | 133 | 18 | 17 | 88 | 107 | 212 | 141 | 203 | 197 | 247 | 46 | 113 | 234 | 67 | 50 | 226 | 125 | 91 |
| 104 | 247 | 107 | 210 | 56 | 2 | 72 | 169 | 214 | 83 | 230 | 241 | 247 | 109 | 247 | 181 | 10 | 83 | 156 | 112 | 8 | 48 | 214 | 107 | 54 | 66 | 175 | 127 |
| 110 | 133 | 73 | 63 | 116 | 57 | 195 | 184 | 172 | 81 | 175 | 128 | 20 | 180 | 176 | 207 | 13 | 215 | 136 | 129 | 122 | 42 | 92 | 137 | 76 | 195 | 27 | 97 |
| 89 | 192 | 245 | 84 | 12 | 130 | 43 | 56 | 126 | 70 | 153 | 173 | 32 | 102 | 94 | 194 | 76 | 41 | 46 | 19 | 207 | 120 | 253 | 12 | 206 | 256 | 255 | 148 |
| 5 | 123 | 5 | 220 | 154 | 161 | 90 | 170 | 71 | 237 | 72 | 200 | 80 | 174 | 234 | 205 | 181 | 93 | 38 | 23 | 225 | 248 | 177 | 211 | 31 | 239 | 31 | 144 |
| 33 | 162 | 98 | 21 | 104 | 40 | 49 | 227 | 218 | 3 | 180 | 127 | 237 | 18 | 90 | 85 | 154 | 241 | 88 | 192 | 217 | 201 | 216 | 228 | 223 | 204 | 44 | 113 |
| 49 | 96 | 122 | 226 | 183 | 75 | 160 | 117 | 167 | 251 | 36 | 219 | 107 | 63 | 213 | 85 | 56 | 228 | 215 | 125 | 121 | 74 | 39 | 176 | 182 | 202 | 67 | 75 |
| 110 | 1 | 211 | 209 | 24 | 238 | 156 | 72 | 34 | 136 | 106 | 250 | 126 | 20 | 194 | 201 | 214 | 140 | 171 | 198 | 173 | 126 | 253 | 204 | 175 | 213 | 53 | 245 |
| 21 | 216 | 128 | 18 | 64 | 178 | 124 | 236 | 237 | 123 | 202 | 230 | 173 | 183 | 152 | 10 | 38 | 189 | 79 | 107 | 131 | 4 | 17 | 168 | 84 | 41 | 163 | 242 |
| 150 | 44 | 30 | 91 | 132 | 224 | 171 | 59 | 73 | 104 | 143 | 112 | 80 | 90 | 224 | 252 | 68 | 84 | 69 | 39 | 203 | 29 | 245 | 103 | 116 | 92 | 45 | 63 |
| 170 | 46 | 236 | 102 | 183 | 168 | 115 | 103 | 80 | 255 | 219 | 59 | 89 | 241 | 189 | 252 | 183 | 181 | 84 | 3 | 85 | 50 | 232 | 117 | 98 | 217 | 221 | 194 |
| 29 | 18 | 125 | 206 | 222 | 179 | 86 | 127 | 204 | 46 | 192 | 24 | 180 | 57 | 152 | 162 | 22 | 9 | 234 | 109 | 130 | 216 | 172 | 186 | 135 | 152 | 153 | 132 |
| 188 | 196 | 75 | 234 | 136 | 191 | 187 | 88 | 49 | 184 | 167 | 240 | 151 | 176 | 228 | 98 | 145 | 191 | 97 | 225 | 83 | 47 | 166 | 102 | 78 | 90 | 180 | 247 |
| 206 | 118 | 32 | 250 | 216 | 226 | 108 | 104 | 244 | 135 | 108 | 66 | 31 | 243 | 248 | 208 | 160 | 68 | 116 | 122 | 125 | 177 | 45 | 164 | 75 | 99 | 23 | 84 |
| 24 | 34 | 80 | 70 | 37 | 48 | 209 | 220 | 159 | 130 | 195 | 122 | 150 | 218 | 173 | 91 | 40 | 26 | 75 | 54 | 231 | 49 | 160 | 9 | 172 | 163 | 13 | 137 |
| 57 | 52 | 17 | 209 | 197 | 217 | 233 | 210 | 178 | 199 | 31 | 29 | 154 | 116 | 10 | 210 | 118 | 9 | 18 | 27 | 52 | 256 | 110 | 55 | 55 | 27 | 251 | 234 |
| 203 | 171 | 185 | 240 | 62 | 117 | 49 | 93 | 121 | 66 | 170 | 230 | 160 | 24 | 16 | 248 | 104 | 8 | 157 | 117 | 157 | 238 | 99 | 225 | 115 | 85 | 253 | 235 |
| 53 | 203 | 100 | 33 | 43 | 49 | 141 | 16 | 49 | 241 | 126 | 28 | 21 | 234 | 247 | 87 | 56 | 228 | 179 | 218 | 94 | 190 | 96 | 35 | 191 | 65 | 219 | 205 |
| 85 | 230 | 66 | 223 | 20 | 155 | 237 | 19 | 31 | 35 | 19 | 203 | 143 | 111 | 236 | 97 | 53 | 230 | 83 | 149 | 79 | 90 | 39 | 51 | 201 | 32 | 107 | 55 |
| 2 | 186 | 64 | 127 | 42 | 93 | 231 | 250 | 61 | 227 | 180 | 16 | 207 | 44 | 29 | 209 | 131 | 200 | 28 | 49 | 202 | 50 | 108 | 51 | 20 | 96 | 222 | [illegible] |

Ryerson University

# HIGHER LEVEL FEATURES

# WHY?

- Labeled data is expensive
- Abundance of unlabeled data
- Less restrictive data requirements

# METHODS

# TWO STAGES

1. Learn representation with unlabeled data
2. Apply to labeled data for classification

# LABELED TRAINING DATA

A set of *m* training examples with:

- *x* as *n*-dimension feature vectors
- *y* as corresponding labels *{1 .. C}*
- *l* indicating a labeled example

$$\left\{ (x_l^{(1)}, y^{(1)}) \ldots \ (x_l^{(m)}, y^{(m)}) \right\} \in R^n$$

Raina et al. (2007)

Ryerson University

## UNLABELED TRAINING DATA

A set of *k* training examples with:

- *x* as *n*-dimension feature vectors
- *u* indicating an unlabeled example

$$\{x_u^{(1)} \dots \; x_u^{(k)}\} \in R^n$$

Raina et al. (2007)

Ryerson University

# LEARNING BASES (BASIC ELEMENTS)

Optimize *aj* weights and *bj* bases to:

1. Reconstruct *xu* as weighted linear combo of bases
2. Encourage *aj* to be sparse (mostly zero)

$$minimize_{b,a} \sum \left\| x_u^{(i)} - \sum_j a_j^{(i)} b_j \right\|_2^2 + \beta \left\| a^{(i)} \right\|_1$$

$$\text{such that } \|b_j\|_2 \leq 1, \forall j \in 1, \ldots s$$

Raina et al. (2007)

# LEARNING FEATURES

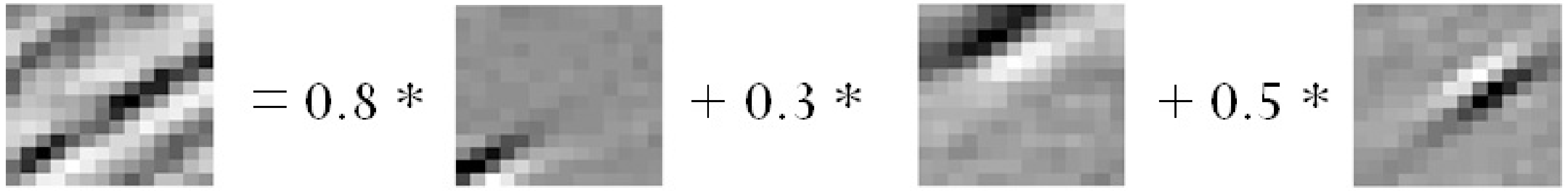
Using *bj*, compute sparse features from labeled data as input to supervised algorithms

$$\hat{a}(x_l^{(i)})) = argmin_a \left\| x_l^{(i)} - \sum_j a_j^{(i)} b_j \right\|_2^2 + \beta \left\| a^{(i)} \right\|_1$$

Raina et al. (2007)

Ryerson University

# SPARSE FEATURES EXAMPLE




$$x \quad = \quad 0.8 * \; b_{87} \quad + \; 0.3 * \; b_{376} \quad + \; 0.5 * \; b_{411}$$

Raina et al. (2007)

# RESULTS

## CLASSIFICATION EXPERIMENTS

- Principal Component Analysis (PCA) vs Raw vs Sparse Coding (SC) Features
- Support Vector Machine (SVM)
- Gaussian Discriminant Analysis (GDA)

## HANDWRITTEN DIGITS AND ENGLISH CHARACTERS

- Improvements when SC used with Raw for characters
- SC did not perform as well alone for characters
- SC generally performed better for digits
- Improvements ranged from ~1-7%

Ryerson University

## REUTERS WEBPAGES AND ARTICLES

- SC generally performed well for webpages and articles
- Improvements ranged from ~8-21%

# KERNELS (SIMILARITY FUNCTIONS)

- Compared against linear, polynomials, Radial Basis Function (RBF)
- Outperforms standard kernel choices above
- Improvements ranged from ~6-13%

Ryerson University

# DISCUSSION AND CONCLUSION

# DISCUSSION

- Can be applied to different domains
- Can have more basis vectors *bj* than n-dimensions
- Unlabeled data must still have some structure
- Other algorithms can be modified for self-taught learning

# CONCLUSION

- Find higher level representations of patterns
- Use of inexpensive unlabeled data
- Self-taught learning as a machine learning framework

# REFERENCES


- Raina, R., Battle, A., Lee, H., Packer, B., & Ng, A. Y. (2007). Self-taught learning. Proceedings of the 24th international conference on Machine learning - ICML '07. doi:10.1145/1273496.1273592