From Digital Images and Text to Vectors

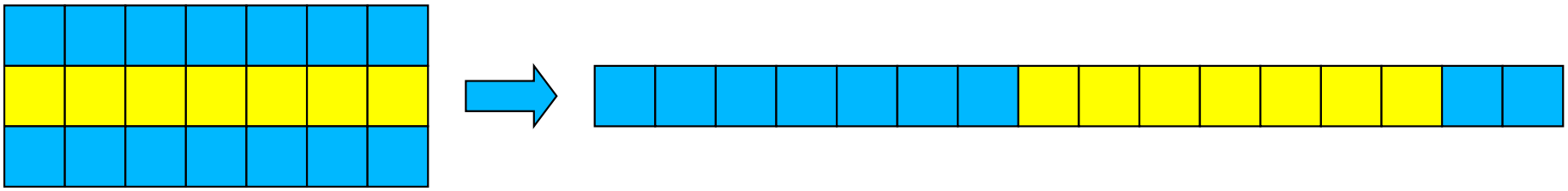
Pascal Fua
IC-CVLab

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Black and White Images

136	134	161	159	163	168	171	173	173	171	166	159	157	155
152	145	136	130	151	149	151	154	158	161	163	163	159	151
145	149	149	145	140	133	145	143	145	145	145	146	148	148
148	143	141	145	145	145	141	136	136	135	135	136	135	133
131	131	129	129	133	136	140	142	142	138	130	128	126	120
115	111	108	106	106	110	120	130	137	142	144	141	129	123
117	109	098	094	094	094	100	110	125	136	141	147	147	145
136	124	116	105	096	096	100	107	116	131	141	147	150	152
152	152	137	124	113	108	105	108	117	129	139	150	157	159
159	157	157	159	135	121	120	120	121	127	136	147	158	163
165	165	163	163	163	166	136	131	135	138	140	145	154	163
166	168	170	168	166	168	170	173	145	143	147	148	152	159
168	173	173	175	173	171	170	173	177	178	151	151	153	156
161	170	176	177	177	179	176	174	174	176	177	179	155	157
161	162	168	176	180	180	180	182	180	175	175	178	180	180

- A black and white digital image is simply a 2D array of numbers, typically coded on 8 bits.

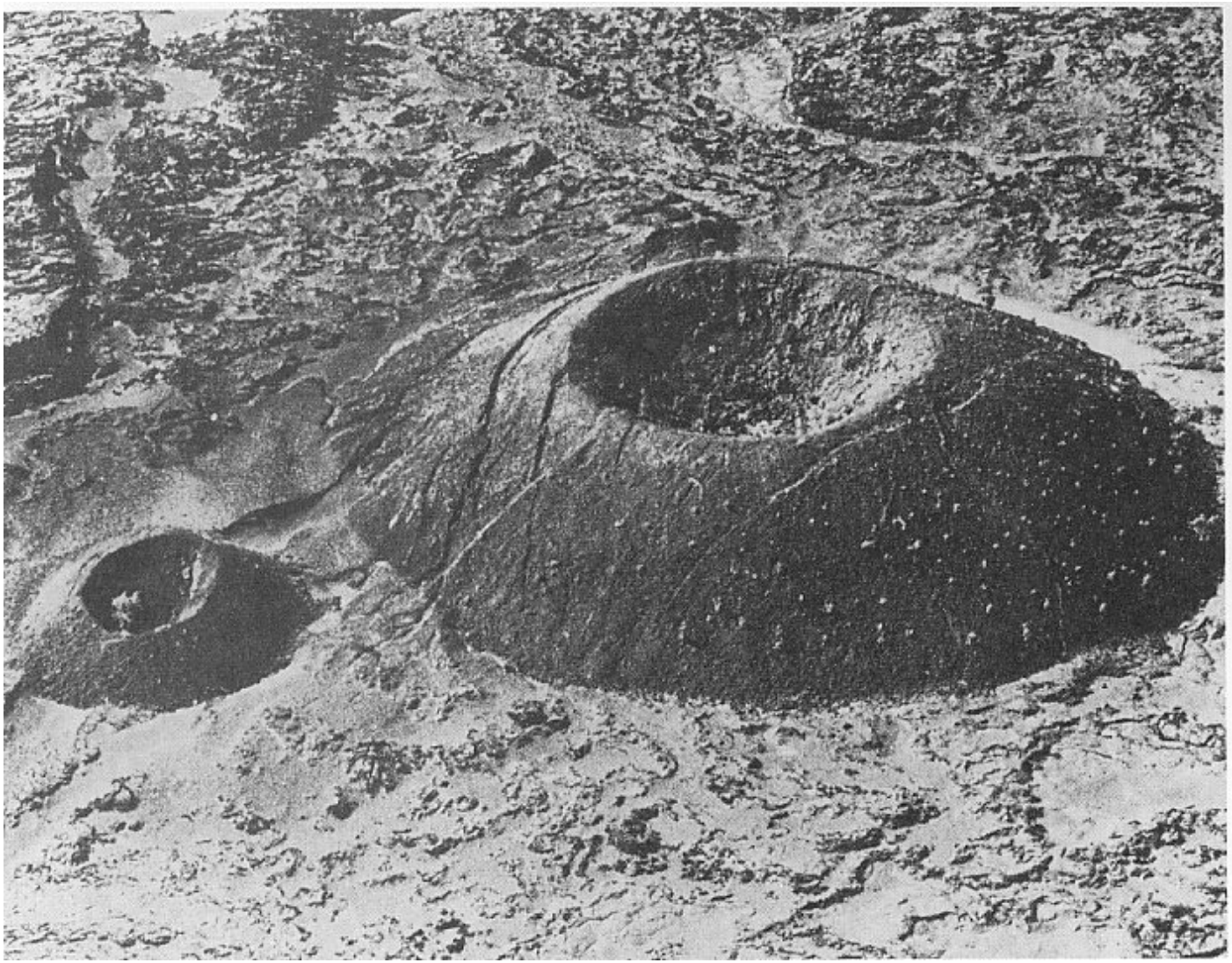


- A MxN image can also be represented as an MN vector.

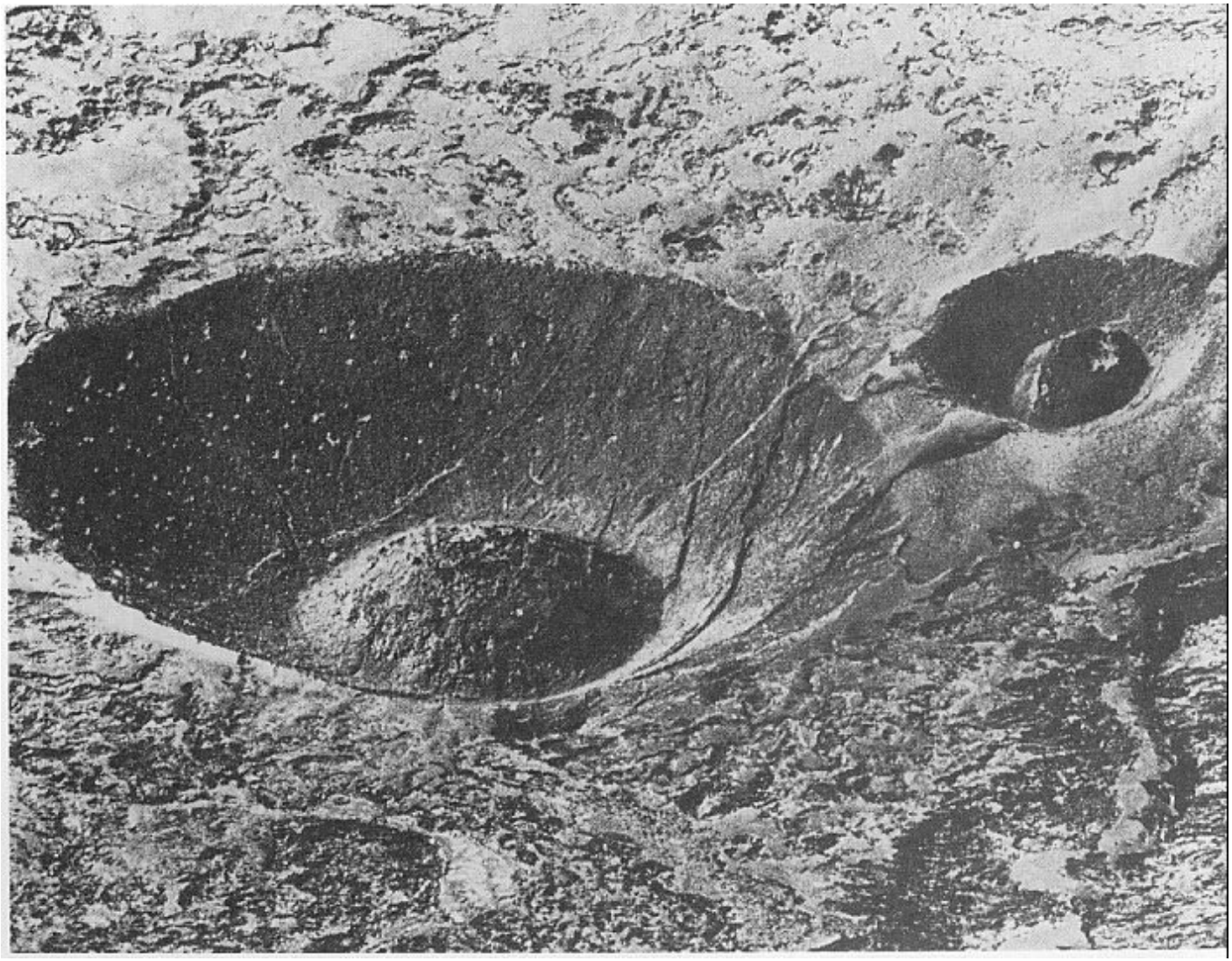
Color Images

A color image is stored as 3 arrays, one for red, one for green, and one for blue.

What DoYou See



What DoYou See



Strong Priors

136 134 161 159 163 168 171 173 173 171 166 159 157 155
152 145 136 130 151 149 151 154 158 161 163 163 159 151
145 149 149 145 140 133 145 143 145 145 145 146 148 148
148 143 141 145 145 145 141 136 136 135 135 136 135 133
131 131 129 129 133 136 140 142 142 138 130 128 126 120
115 111 108 106 106 110 120 130 137 142 144 141 129 123
117 109 098 094 094 094 100 110 125 136 141 147 147 145
136 124 116 105 096 096 100 107 116 131 141 147 150 152
152 152 137 124 113 108 105 108 117 129 139 150 157 159
159 157 157 159 135 121 120 120 121 127 136 147 158 163
165 165 163 163 163 166 136 131 135 138 140 145 154 163
166 168 170 168 166 168 170 173 145 143 147 148 152 159
168 173 173 175 173 171 170 173 177 178 151 151 153 156
161 170 176 177 177 179 176 174 174 176 177 179 155 157
161 162 168 176 180 180 180 182 180 175 175 178 180 180

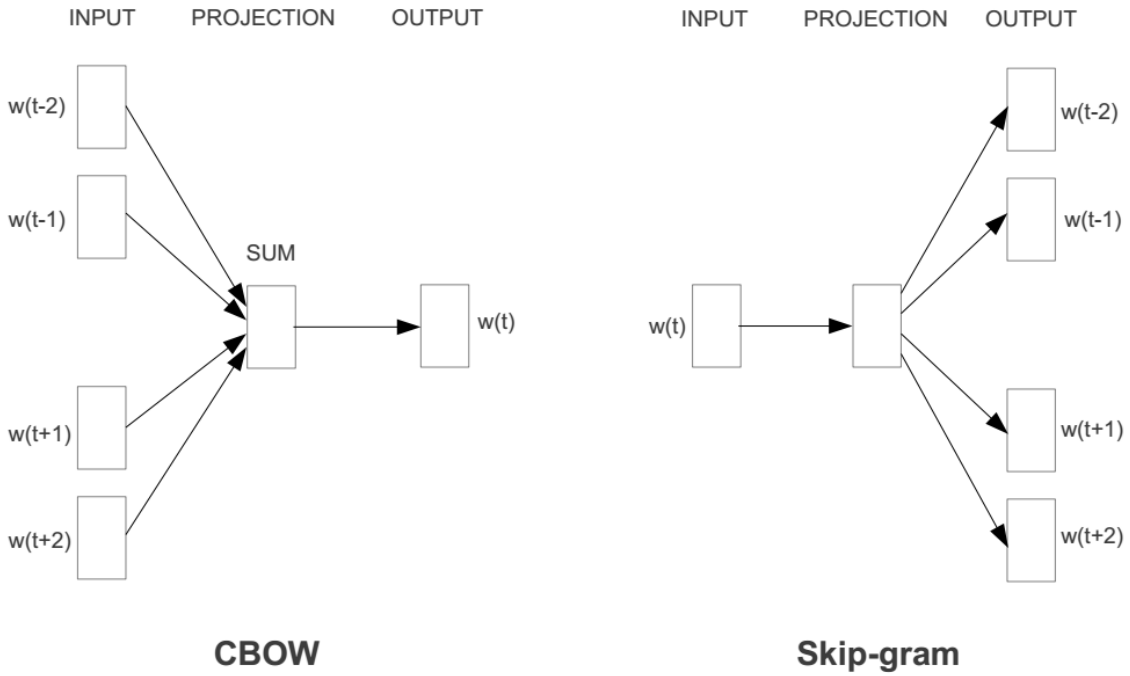
Converting Words to Vectors

- How similar is **pizza** to **pasta**?
- How related is **pizza** to **Italy**?

- Representing words as vectors allows for easy computation of similarity.
- Makes it possible to use the Machine Learning techniques we have discussed.
- Exploit the theory is that similar words occur in similar context.

Harris, Word'54, Firth; Studies in Linguistic Analysis'57

Words in Context
word2vec



Two basic neural network models:

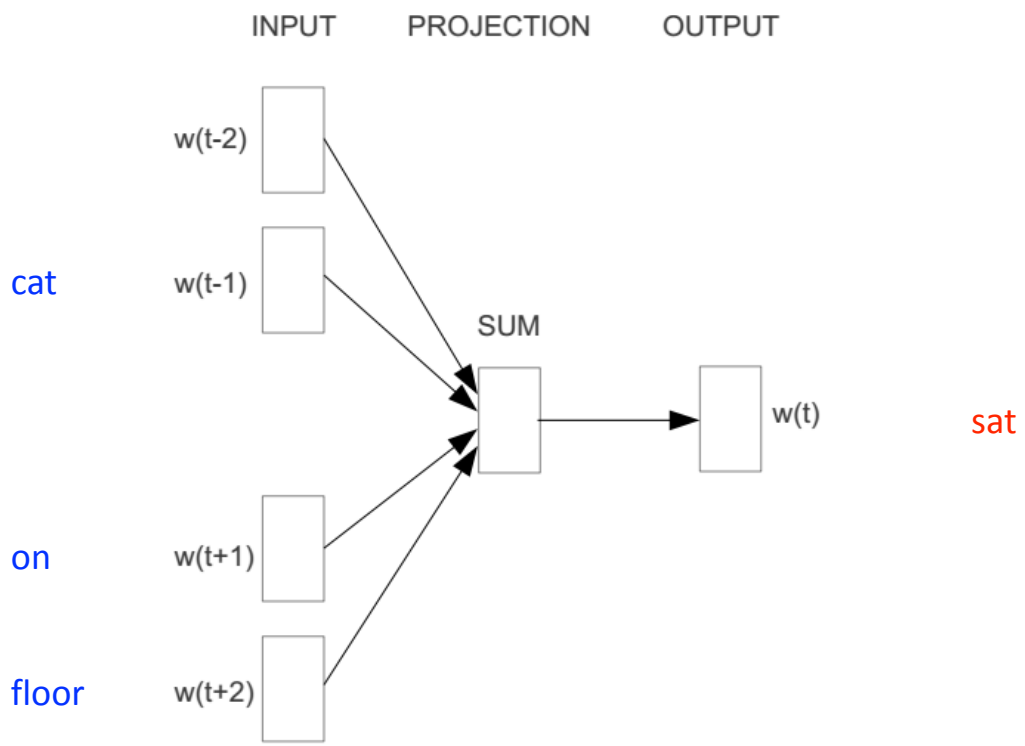
- Continuous Bag of Word (CBOW). Use a window of words to predict the middle one.
- Skip-gram (SG). Use a word to predict the surrounding ones in a window.

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Mikolov et al., ArXiv'13

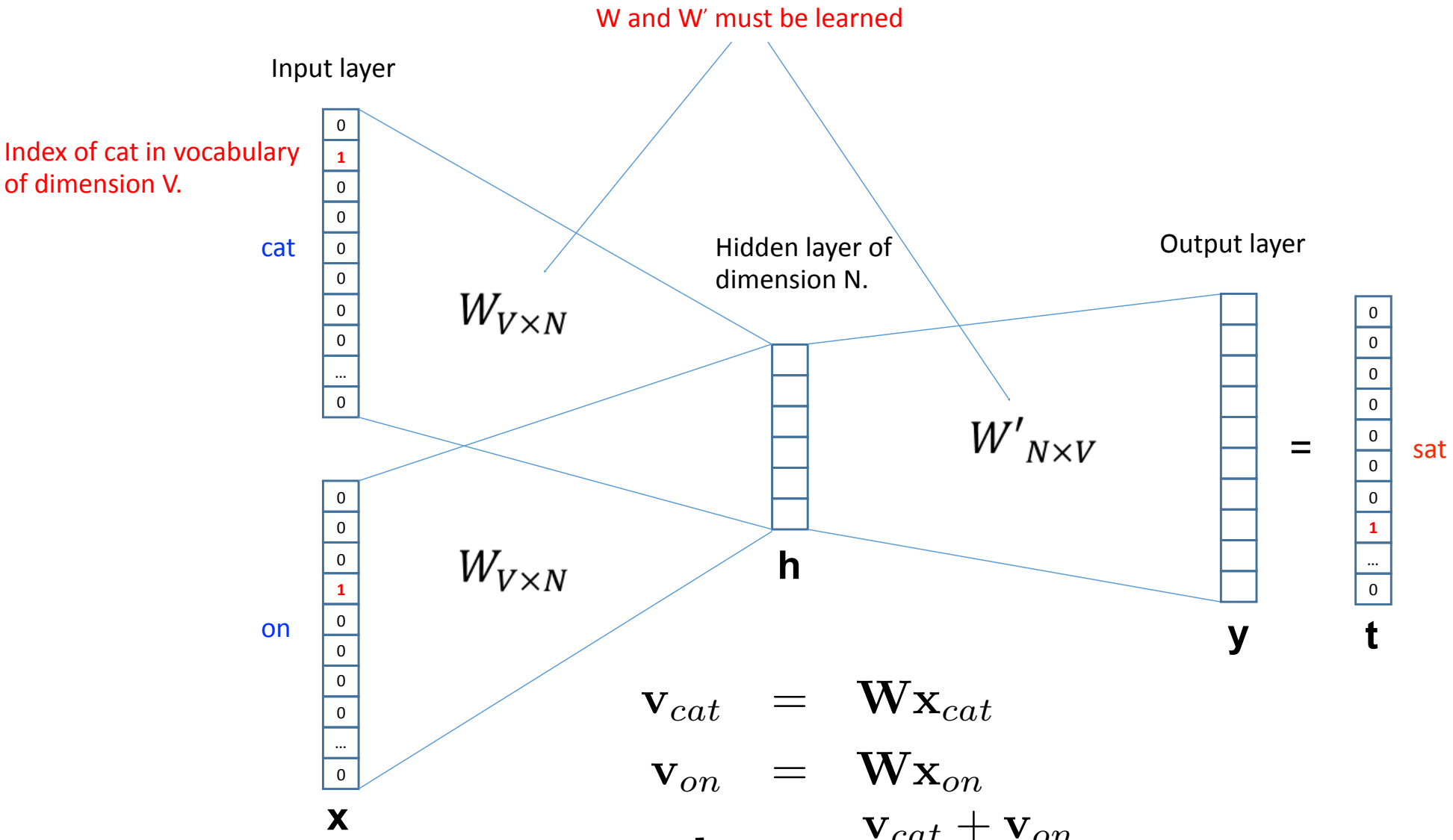
Continuous Bag of Words

The cat sat on the floor.



sat should be predicted from the words around it.

Window of Size 2



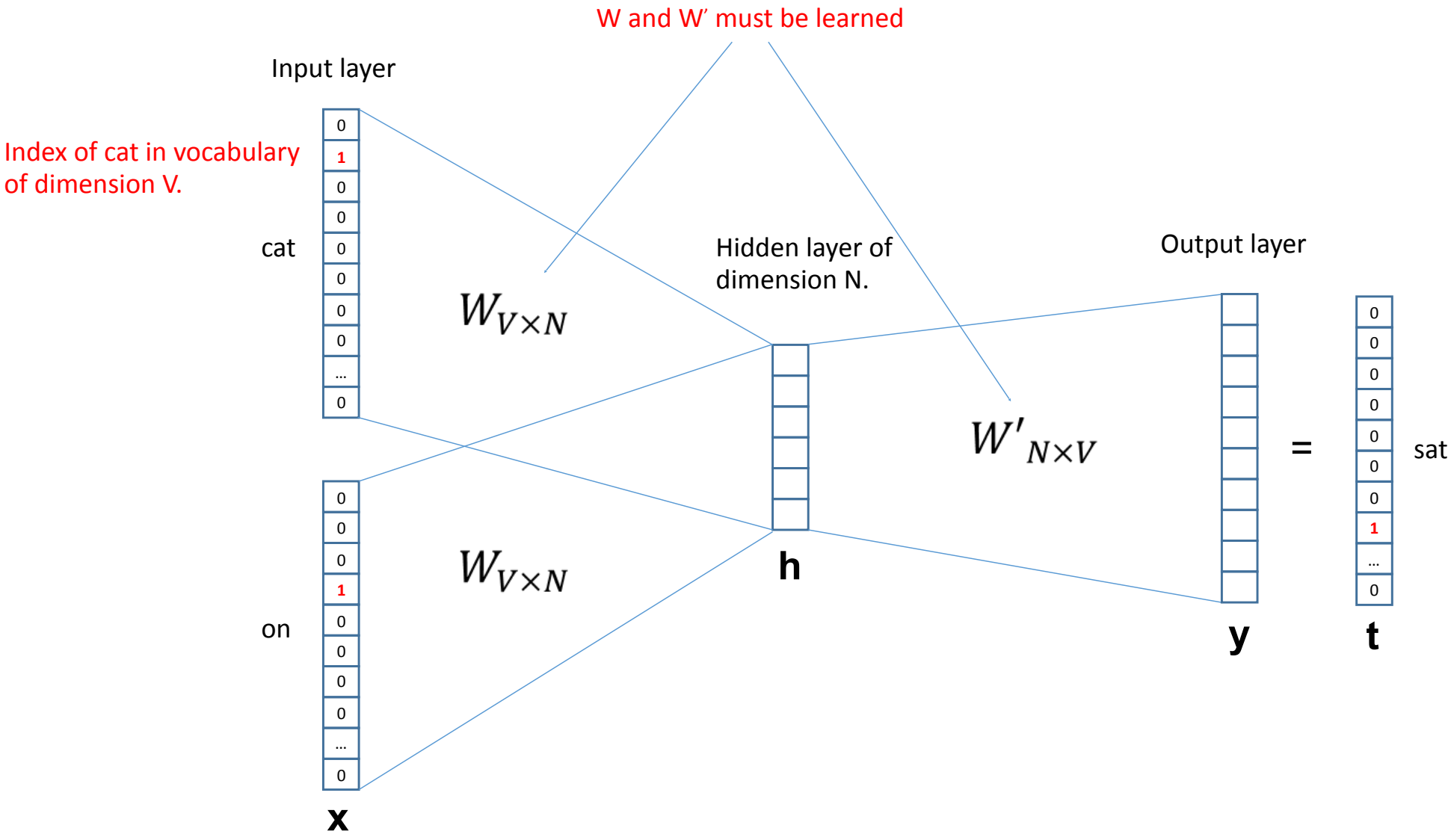
$$\mathbf{v}_{cat} = \mathbf{W}\mathbf{x}_{cat}$$

$$\mathbf{v}_{on} = \mathbf{W}\mathbf{x}_{on}$$

$$\mathbf{h} = \frac{\mathbf{v}_{cat} + \mathbf{v}_{on}}{2}$$

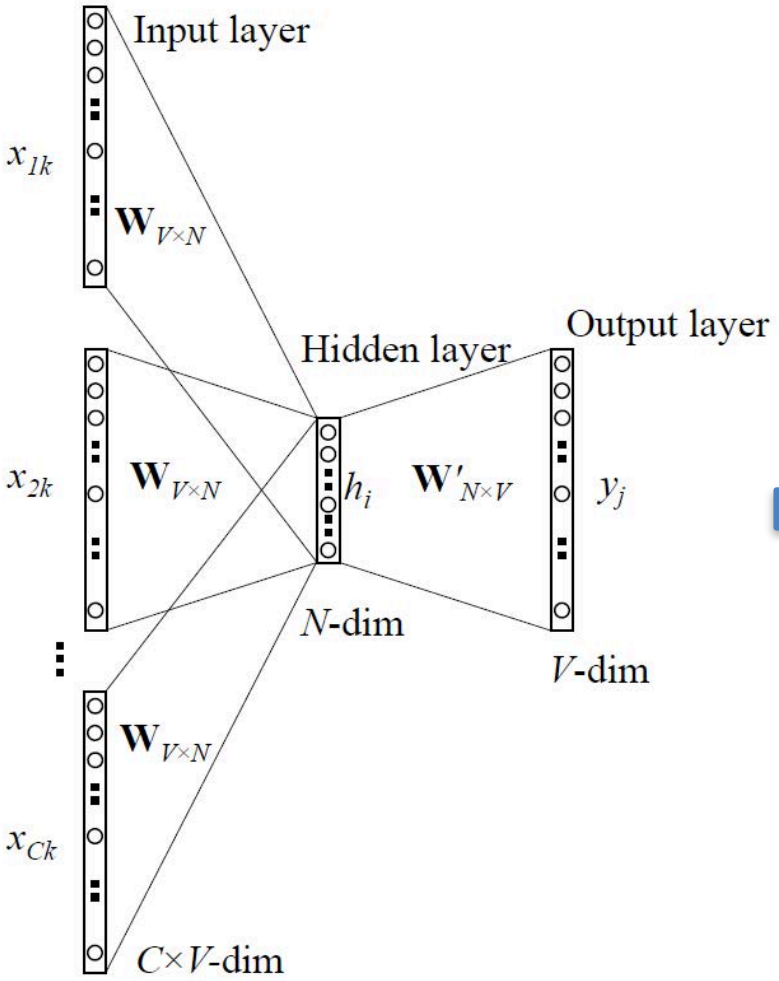
$$\mathbf{y} = \text{softmax}(\mathbf{W}'\mathbf{h})$$

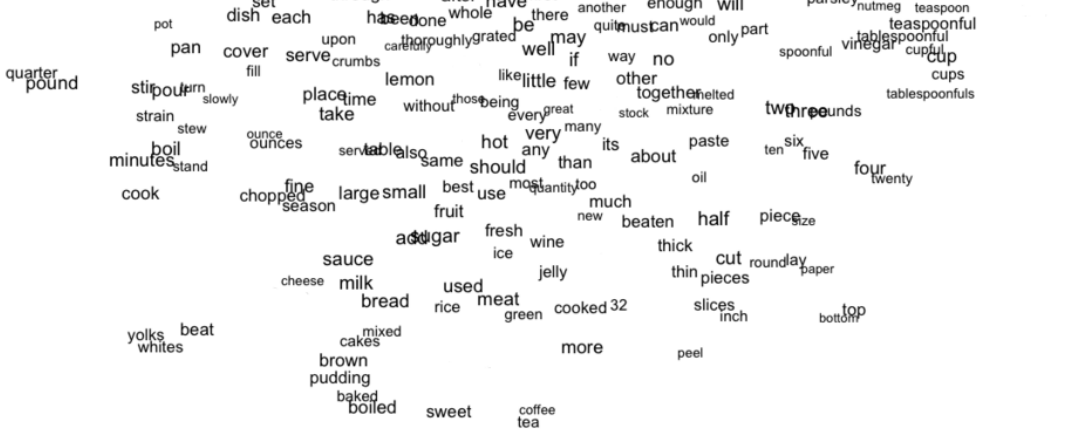
Window of Size 2



Once the training is complete, the W matrices associate to each word a vector of dimension N.

Larger Windows





Once the training is complete, the W matrices associate to each word a vector of dimension N.