



### 【変数定義】

ボキャブラリ数（語彙数）：vocab\_size = 9  
ウィンドウサイズ: window = 1  
中間層のユニット数: hidden = 3  
入力ベクトル:  $x$   
予測結果:  $y$   
教師データ:  $t$

### 【行列定義】

#### 入力ベクトル： $x$

行数=vocab\_size\*(window\*2)=18  
=> 18行1列

#### 入力層・中間層の重み行列： $W_{in}$

行数=hidden = 3  
列数= vocab\_size\*window\*2 = 18  
=> 3行18列（54個の重み）

#### 中間層・出力層の重み行列： $W_{out}$

行数=vocab\_size = 9  
列数=hidden = 3  
=> 9行3列（27個の重み）

### 【推論の流れ】

1.  $W_{in}, W_{out}$  を初期化する。
2. サンプル毎に入力ベクトル $x$ と教師データ $t$ を用意。
3. 中間層の出力を算出。  
 $hidden = activation(W_{in} * x) = 3 \text{ 行 } 1 \text{ 列}$
4. 出力層の出力を算出。  
 $output = W_{out} * hidden = 9 \text{ 行 } 1 \text{ 列}$
5. 出力をsoftmaxレイヤに適用。)  
左図には記載なし。Softmax で確率を算出、最大確率となる語彙を取得。  
 $y = \arg\_max(\text{softmax}(\text{output}))$