

気持ち重視で 解説する Word2vec

2021/03/15

発表の意図

- Word2vecについて
「単語を推論したら、同じような単語が同じようなベクトルになる」
ということは聞いたことあるけど、
「じゃあなんで推論したら同じような単語が同じようなベクトルになるか」
と言われるとよくわかっていなかった
- ので改めて勉強して「なるほど～」となったので共有したい
- そういう状況の人がいれば今回の発表で分かるようになるはず

他力本願

- 仏のようなスライド
(https://www.ailab.ics.keio.ac.jp/b4_induction_training/docs/nlp/dl_fromscratch2_nlp.pdf) を参考にさせていただく
- この中で登場するCBOWモデルの気持ち部分を補足するためのスライドとして聞いてほしい

One-hotベクトルのところだけ復習

one-hot vector

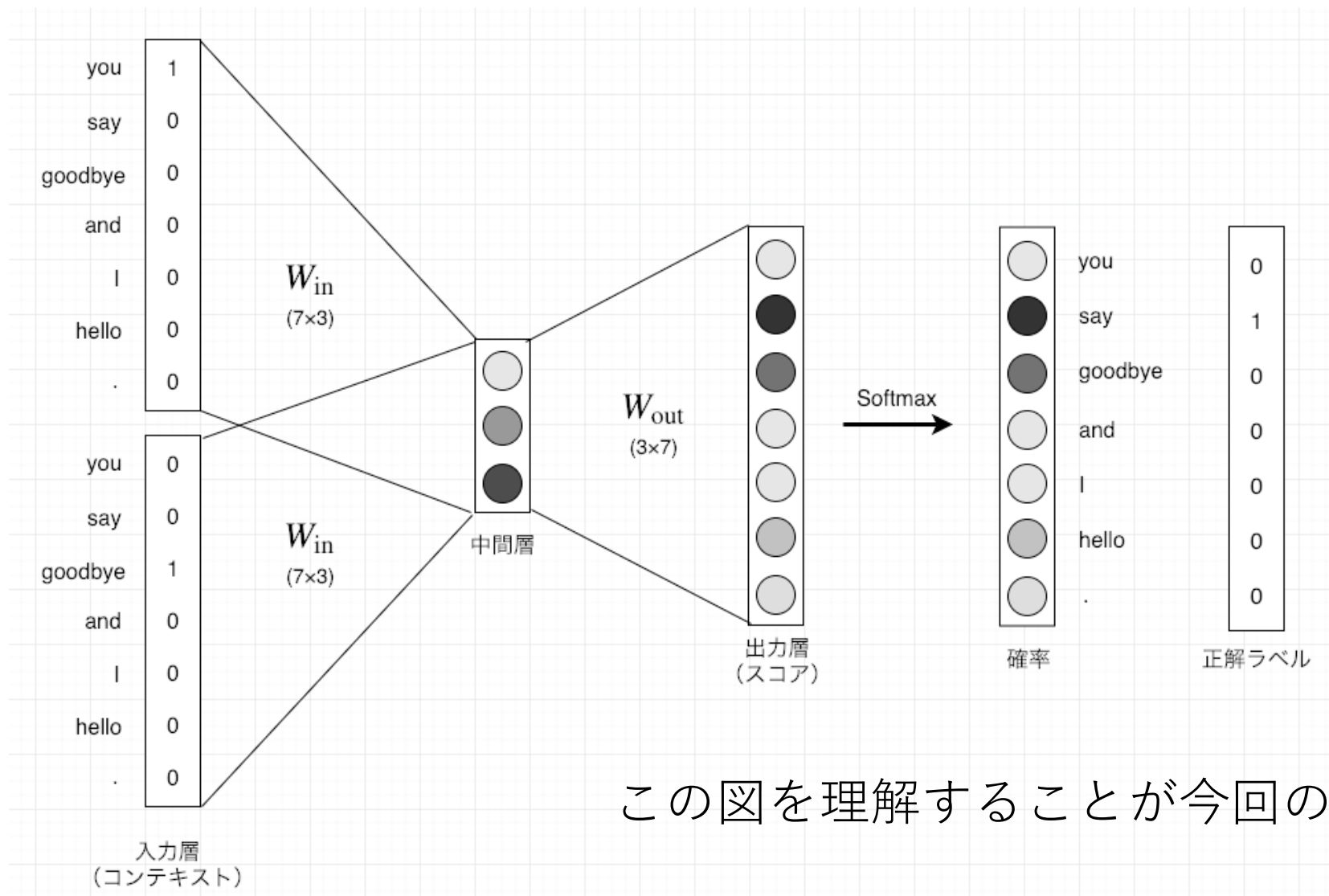
You say goodbye and I say hello.

という1文をコーパスとする(2章と同じ)

単語	単語ID	one-hot vector
you	0	(1,0,0,0,0,0,0)
goodbye	2	(0,0,1,0,0,0,0)

ピリオドを含む8単語を含む文章を全テキストデータ（コーパス）とし、大きさ7の辞書が生成されていることを確認してください

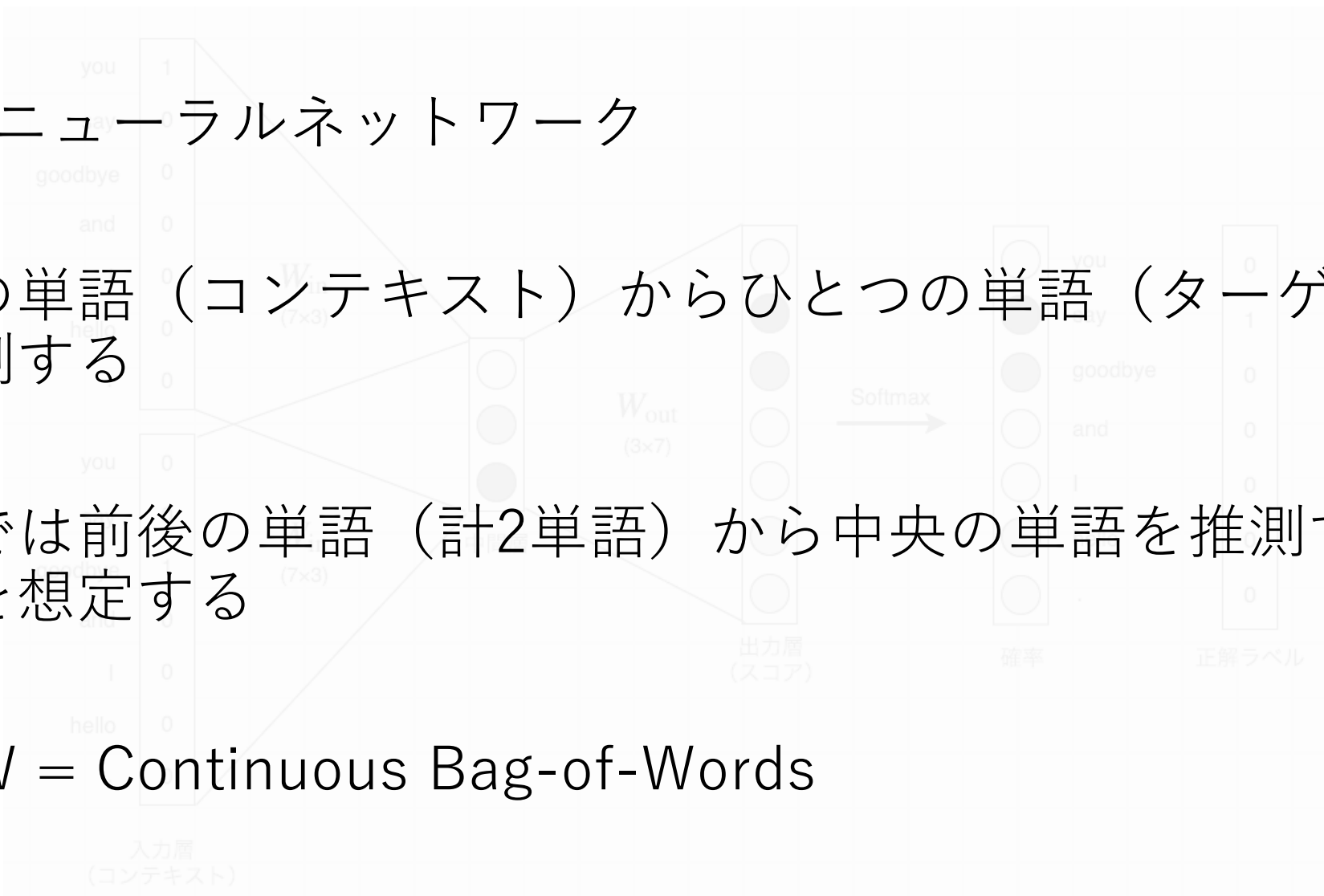
CBOWモデル概要図



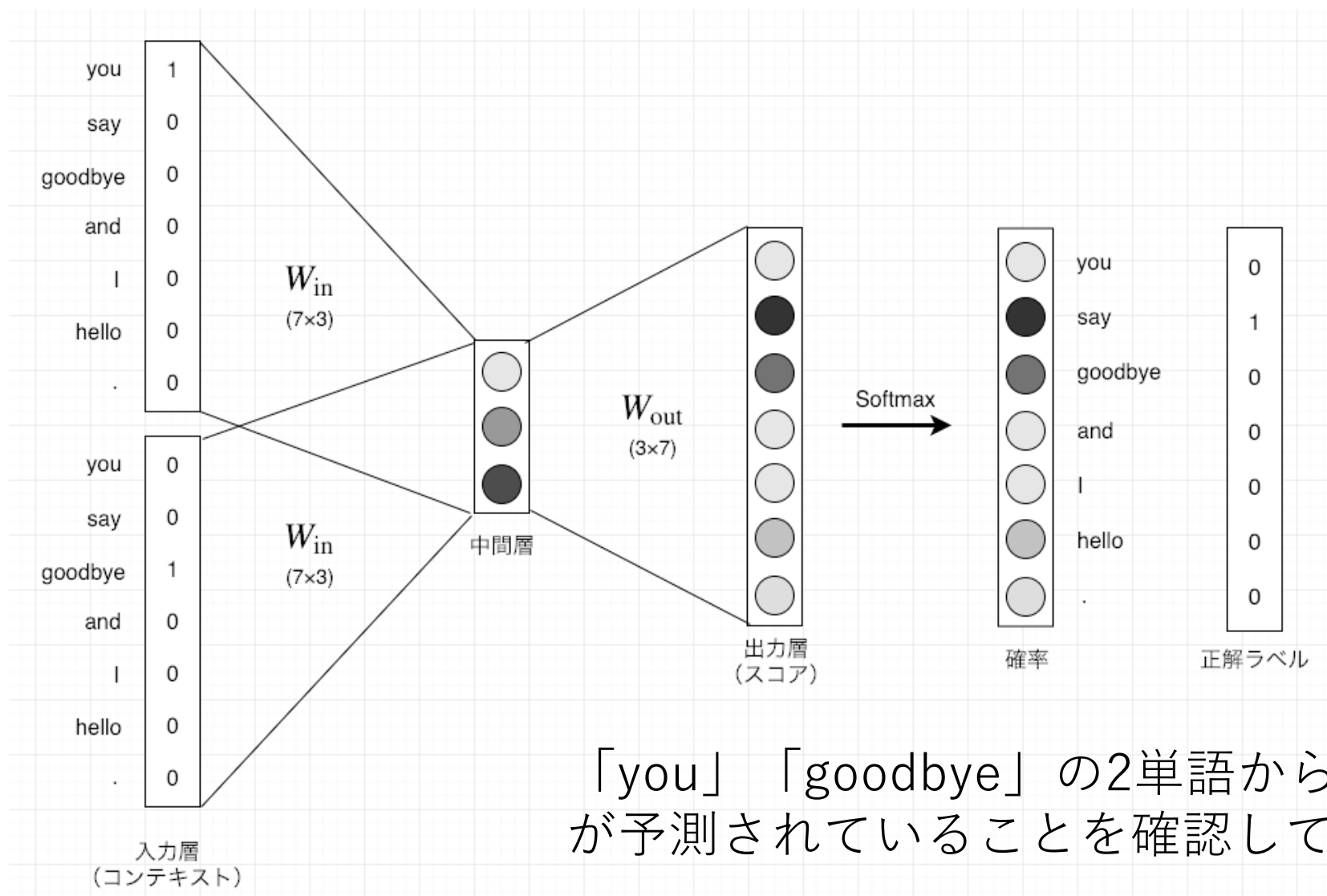
この図を理解することが今回の目標です

CBOWモデルとは

- 2層のニューラルネットワーク
- 周囲の単語（コンテキスト）からひとつの単語（ターゲット）を推測する
- ここでは前後の単語（計2単語）から中央の単語を推測するモデルを想定する
- CBOW = Continuous Bag-of-Words



モデル概要図



モデル構造

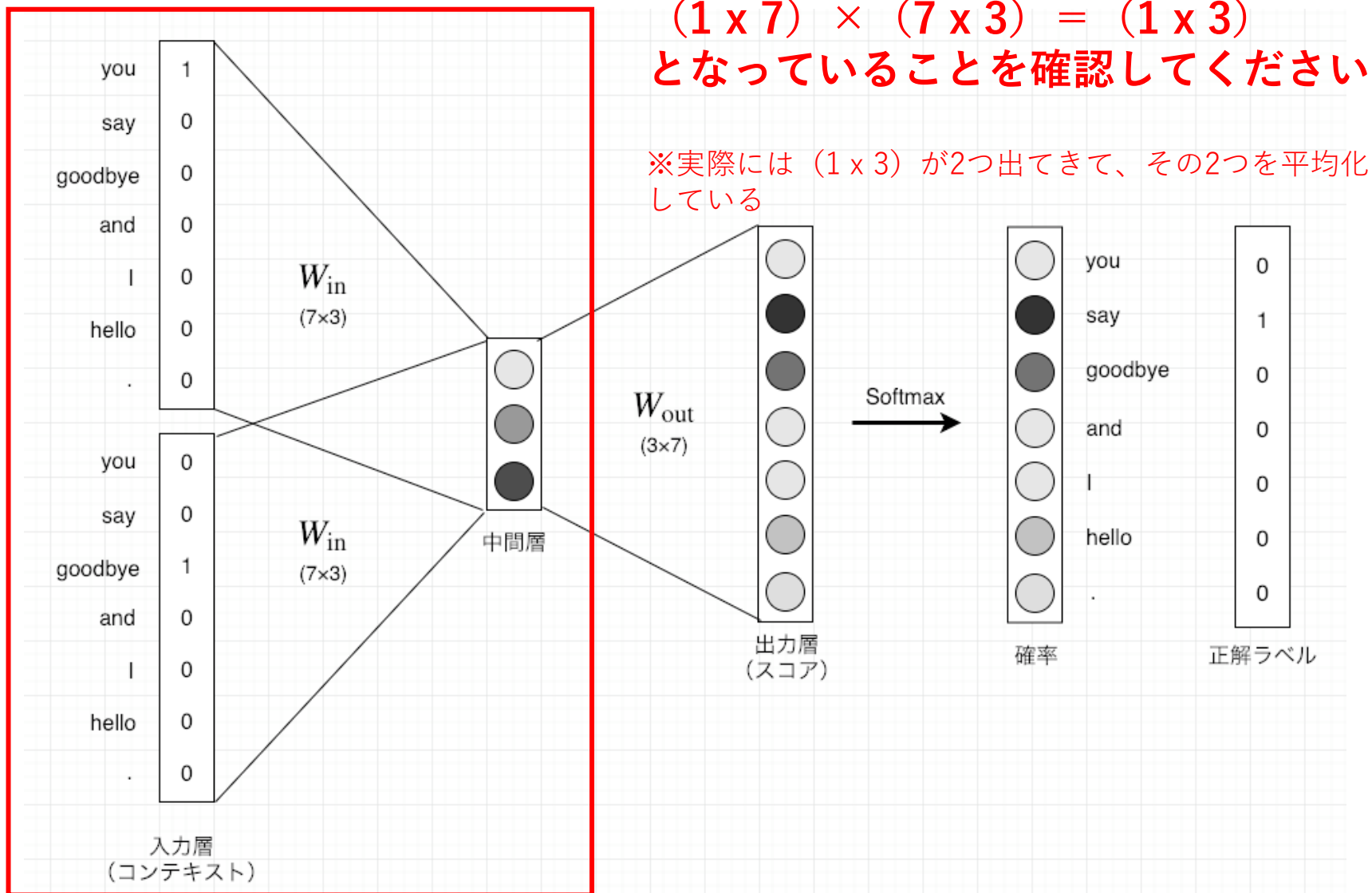
- ① 単語をone-hotベクトルに変換 (1×7)
- ② 1層目の重み (7×3) と掛け合わせる
 - (1×3) の行列が出力される
- ③ 2層目の重み (3×7) と掛け合わせる
 - (1×7) の行列が出力される
- ④ 出力 (1×7) にSoftmaxする
 - (1×7) の $0 \sim 1$ の値を持つ行列が出力される
- ⑤ 損失計算し、1層目と2層目の重みを更新する

※データ中の総単語数：7, 中間層の次元(要素数)：3の場合

CBOWモデルの気持ち解説

- 1層目は**単語変換**の役割
 - CBOWモデルは周囲の単語から中央の単語を予測するモデル
 - 予測にはone-hotベクトルだけだと情報が少なすぎる
 - 予測担当の2層目のために、one-hotベクトルを何かしらに変換する役割**
- 2層目(+Softmax)は**予測**の役割
 - 1層目に変換した「前の単語を変換したもの」と「後ろの単語を変換したもの」を入力とする
 - 中央の単語を予測し、全単語に対して「中央の単語である確率」を算出する

モデル概要図



1層目の解説

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \\ 16 & 17 & 18 \\ 19 & 20 & 21 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

入力
(1 x 7)

1層目の重み w_{in}
(7 x 3)

中間層
(1 x 3)

1層目の解説

1 0 0 0 0 0 0	T		
	1	2	3
	4	5	6
	7	8	9
	10	11	12
	13	14	15
	16	17	18
0	19	20	21

「You」部分

「say」部分

「goodbye」部分

⋮

入力 1層目の重み w_{in}
(1 x 7) (7 x 3)

w_{in} は1行が1単語を表す単語の集合体！！！！

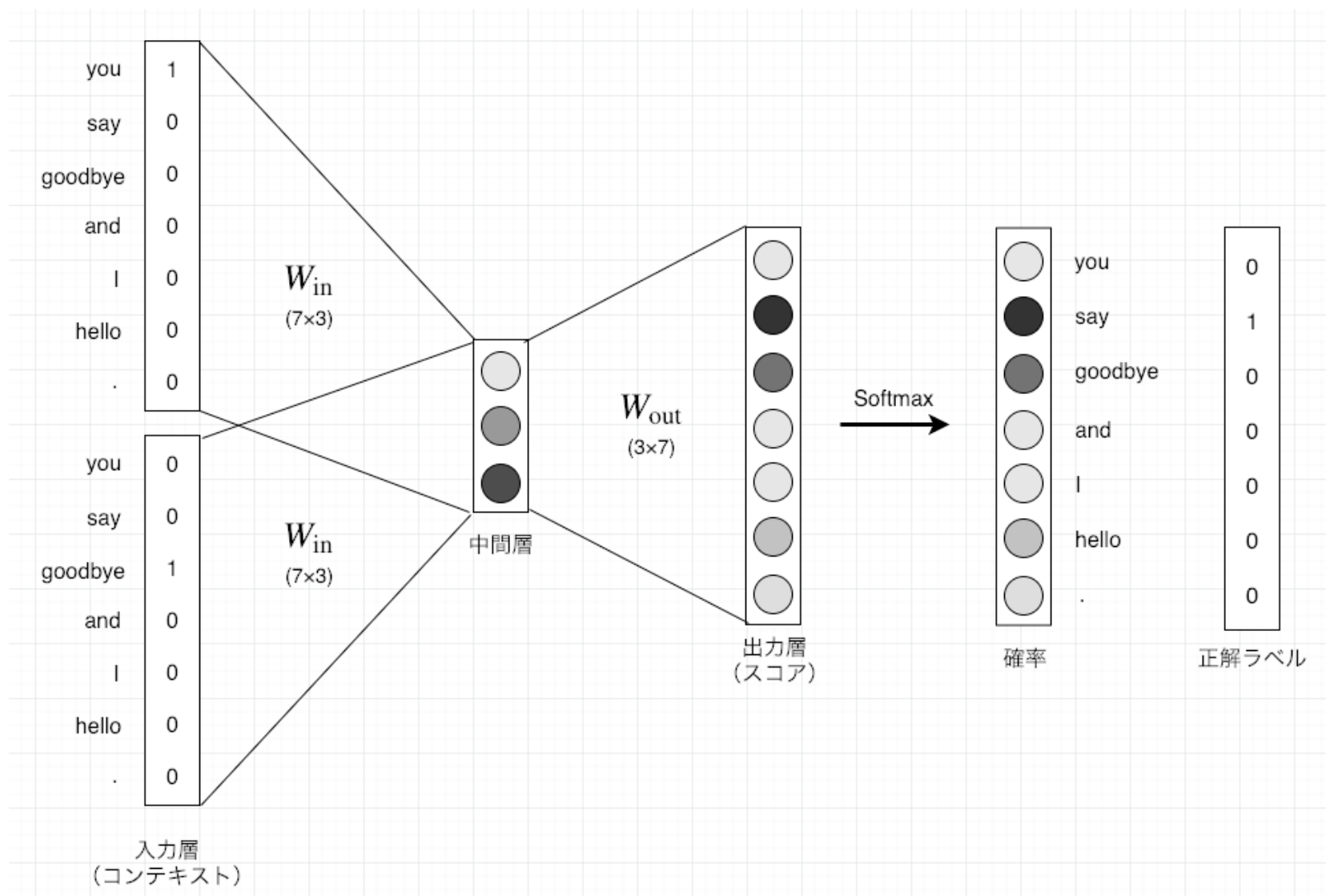
1層目と2層目はそれぞれ何を指すか

- 1層目は予測担当の2層目のために、one-hotベクトルを何かしらに変換して渡す役割
- 「I」「You」の後には同じような単語が来ることが予想されるので、2層目はこれらの単語に対して同じような中間層を受け取れるとありがたい
- それができないと予測がうまくできないので損失の値が大きくなって更新される
- 「予測しやすい中間層を提供する」ことが1層目の目標

1層目と2層目はそれぞれ何を指すか

- 2層目は中間層を受け取って中央の単語を予測する
 - つまり受け取るのは「I」や「You」ではなく謎の(1 x 3)の行列
 - 予測した後Softmax、正解ラベルと照合し、損失を取る
 - 予測が正確にできるようになるまで更新を続ける
-
- 正確な予測のために2層目の予測もうまくなっていくし、1層目の変換もうまくなってくる

CBOWモデル



まとめ

- 1層目は2層目のために「予測しやすい中間層を提供する」よう更新される
- 2層目は中間層を受け取って中央の単語を予測する
- そうすると似た単語が似た値を持つような w_{in} がつくられる
- One-hotベクトルを特徴量に変換する w_{in} が完成した！
- これで単語をベクトル化できるね
- 2層目バイバイ