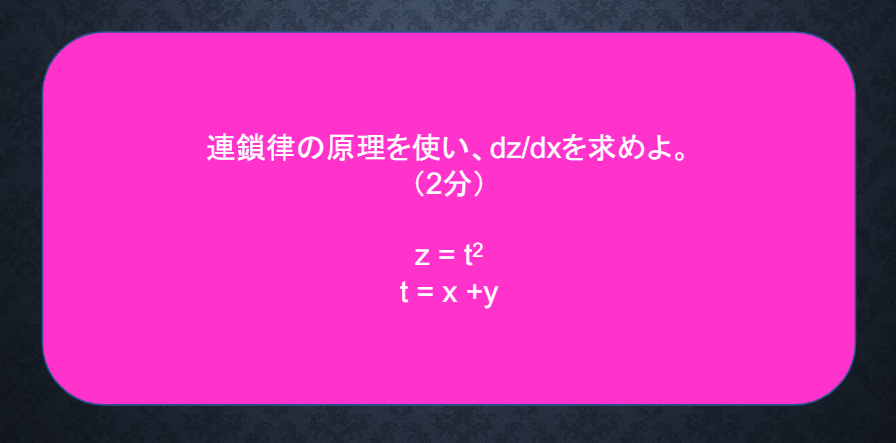
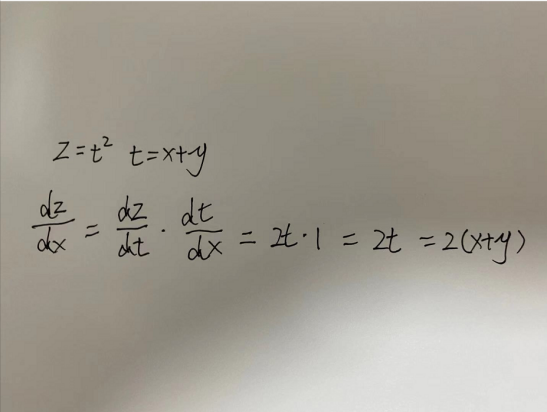
**まとめ要点１：**

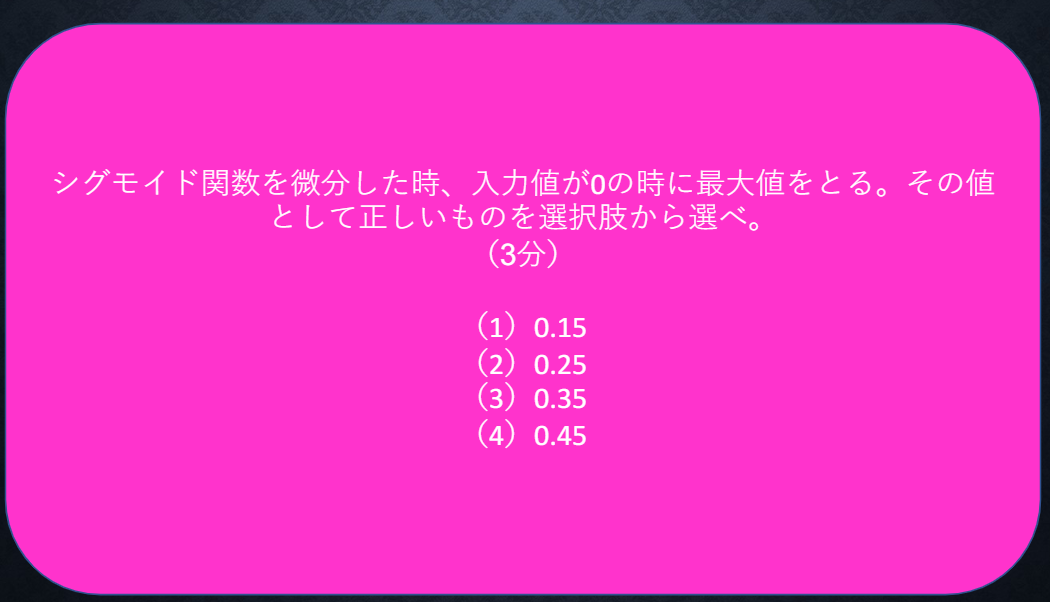
　勾配消失問題について

勾配消失問題は誤差逆伝播法が下位層に進んでいくに連れて、勾配がどんどん緩やかになっていく。そのため、勾配降下法による、更新では下位層のパラメータはほとんど変わらず、訓練は最適値に収束しなくなるという問題。勾配消失問題はsigmoid関数などの活性化関数と関連性は、sigmoid関数は0~1の間で大きな値では出力の変化が微小なため、勾配消失問題を引き起こす事があった。その解決策としてはRELU活性化関数を使用すること。

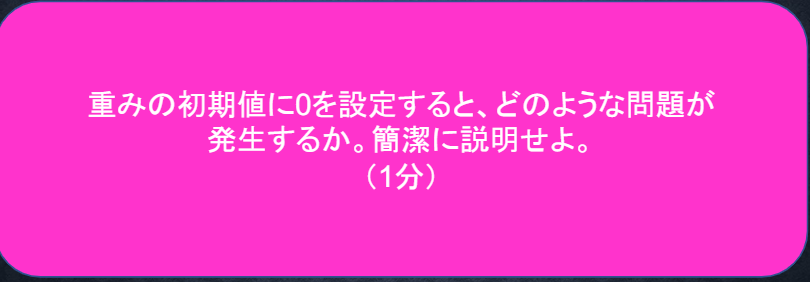
**確認テスト;**



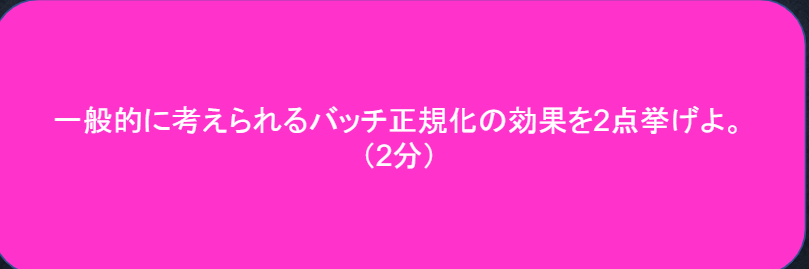
**答え：**



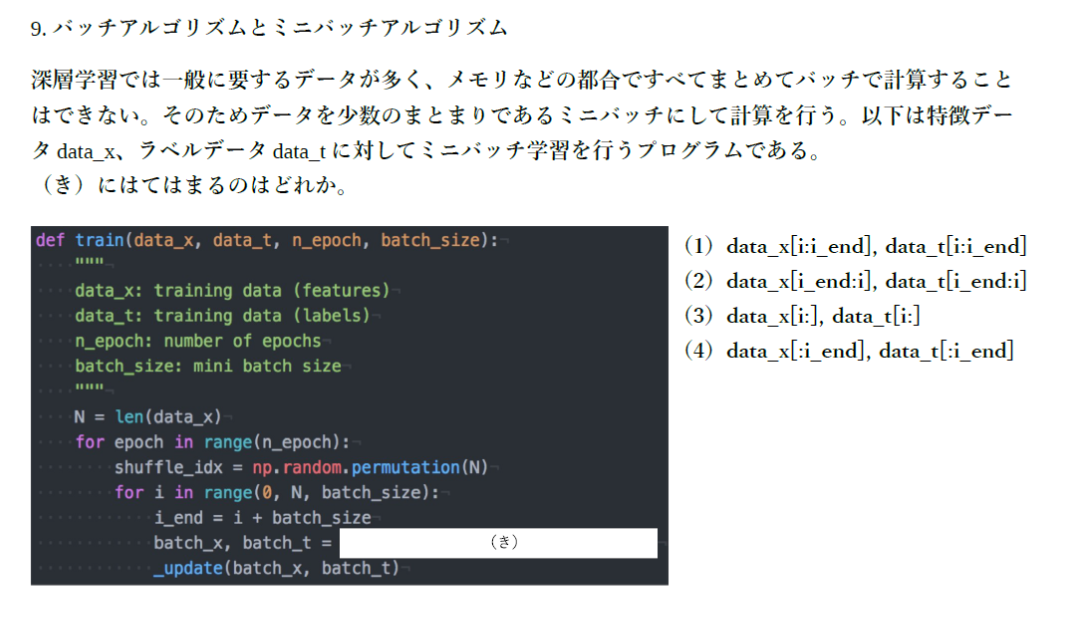
**答え：sigmoid関数を微分した後の式を利用して、０の時、最大値を取ることが出来るので、０を代入して計算すること、0.25、2番を選択することが分かりました。**



**答え**：**全て重みは均一に更新されるため、学習が行えないことになった。**



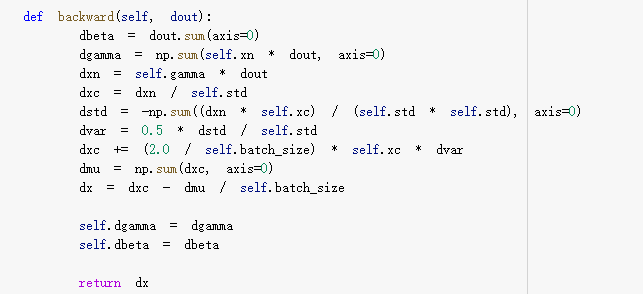
**答え**：**学習を安定化するや計算スビートアップすることができます。**



**答え：バッチサイズだけデータを取り出す処理である、1を選択する。**

**実装確認**





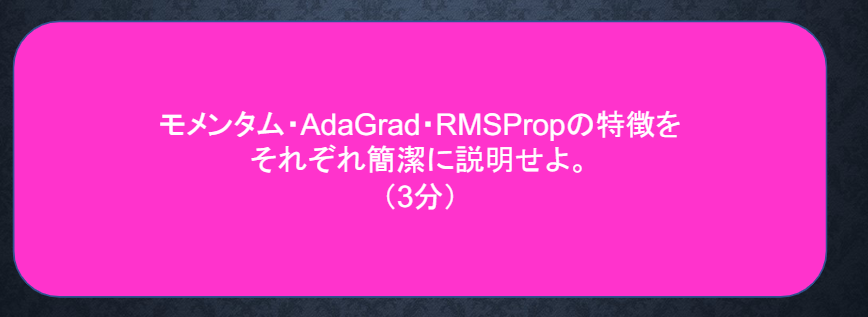


**初期値、バッチサイズなどのハイパーパラメータを調整することで、結果も変化している。**

**まとめ要点2：**

　学習率最適化手法について

学習率を最適するための手法はいくつかがあります。よく使う手法として、モメンタムやAdaGradやRMSPropやAdamなど。モメンタム：誤差をパラメータで微分したものと学習率の積を減算した後、現在の重みに前回の重みを減算した値と慣性の積を加算する。AdaGrad：誤差をパラメータで微分したものと再定義した学習率の積を減算する。RMSProp：誤差をパラメータで微分したものと再定義した学習率の積を減算する。Adam：・モメンタムの過去の勾配の指数関数的減衰平均とRMSPropの過去の勾配の2乗の指数関数的減衰平均を組み合わせての最適化手法。



**答え：まとめに書いているので、ここは省略する。**

**実装確認：**

モメンタム：



AdaGrad：



RSMprop：



Adam：



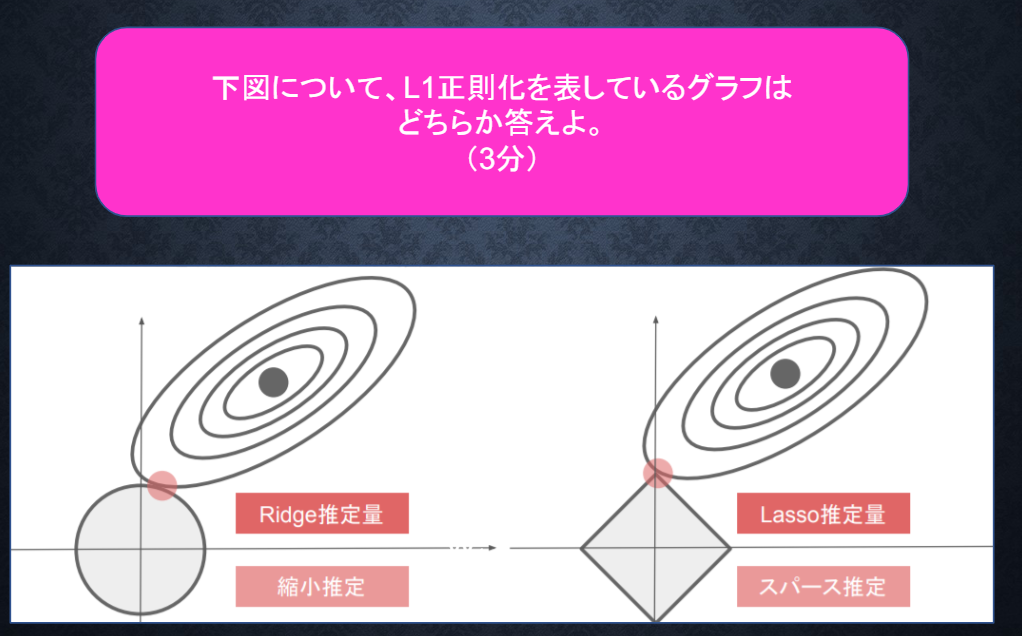
**学習率の変更や重みの初期化パラメータ変更、バッチ正規化などの調整により、結果は変わります。**

**まとめ要点３：**

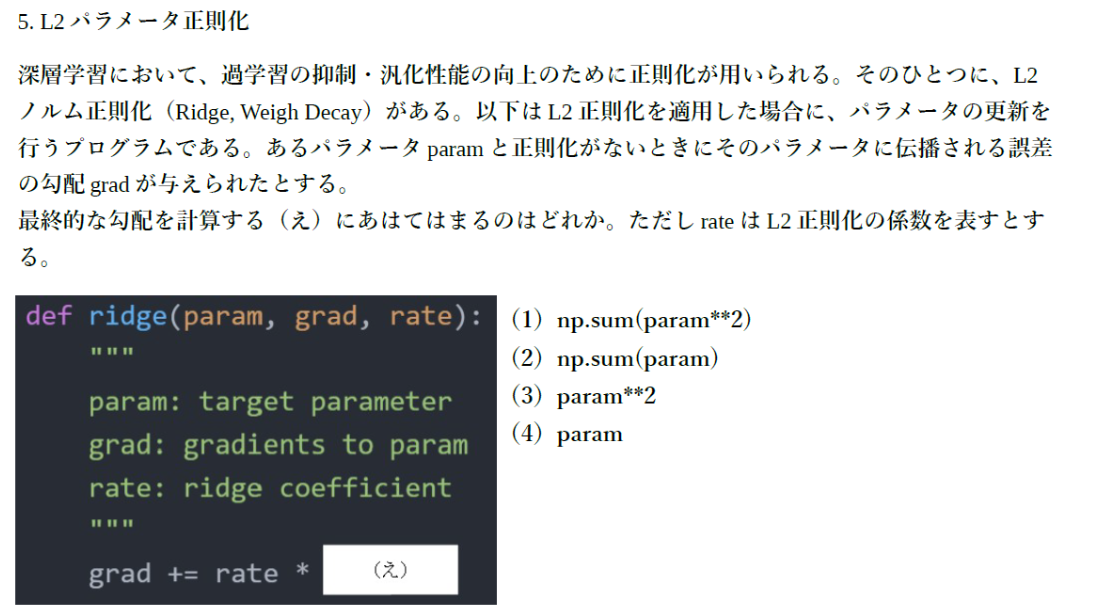
　過学習

過学習の基本概念は既に勉強していたので、説明を省略する。この章は過学習の対処法を紹介している。基本的には、誤差に対して、正則化項を加算することで、重みを抑制する正則化手法(L1、L2)、とランダムにノードを削除して学習させることで過学習、ドロップアウト層の追加して二択あります。

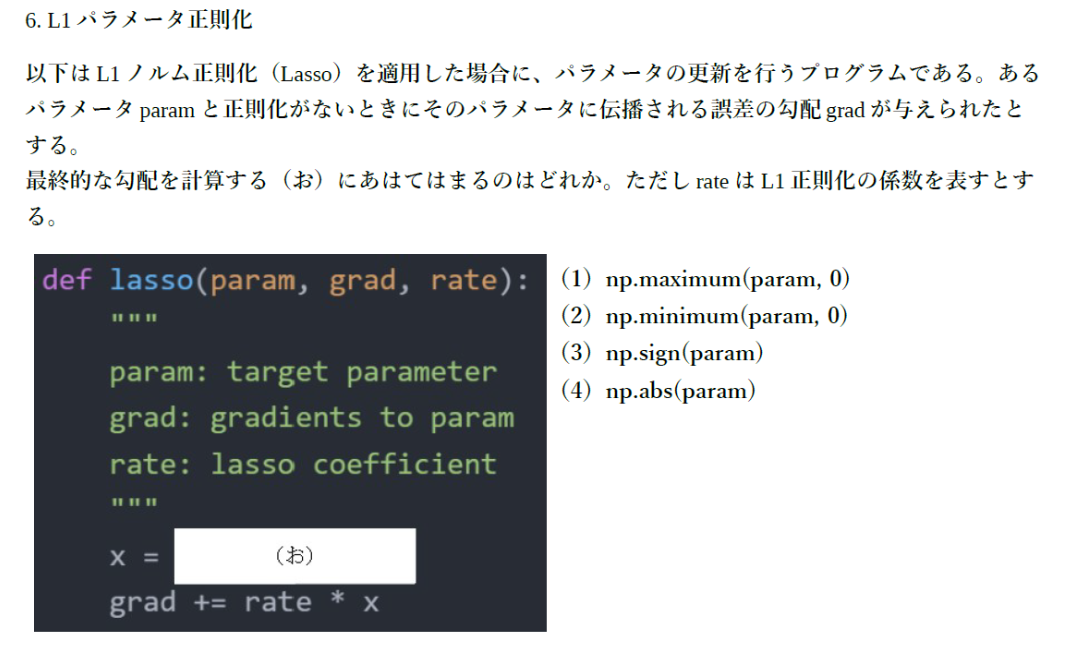
**確認テスト：**



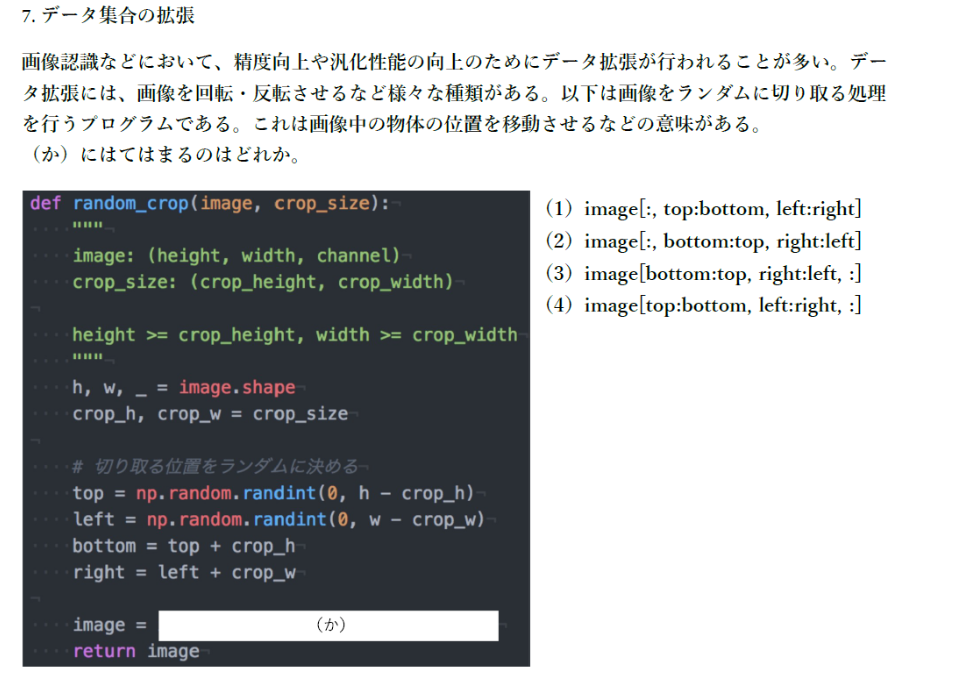
**答え：L1はLasso回帰なので、、右のグラフは正解。**



**答え：Param**



**答え：(３)を選択する、signは符号関数である。**



**答え：imageの形式が(縦幅, 横幅, チャンネル)であるなので、(４)は正解。**

**実装演習：**

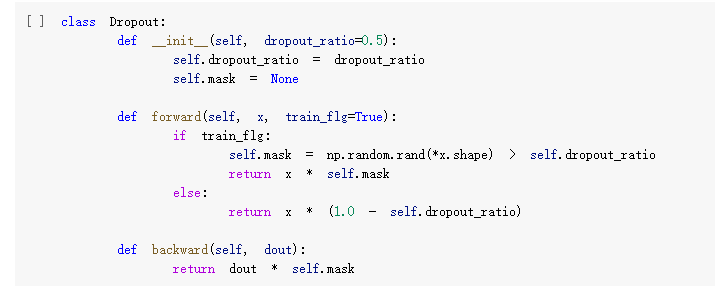
L1：



L2：



Dropout：



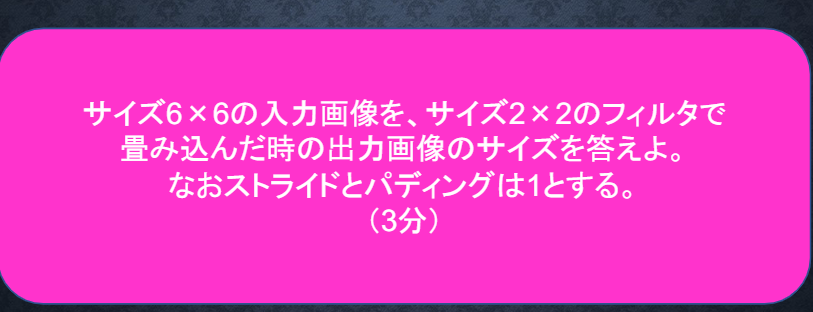
**＊実際に実装する時、正規化やdropoutなどの手法はDLエンジニア(tensorflowなど)に用意されていたので、それを使えば便利かつ間違ないなどのメリットがある。**

**まとめ要点４：**

　畳み込みニューラルネットワークの概念

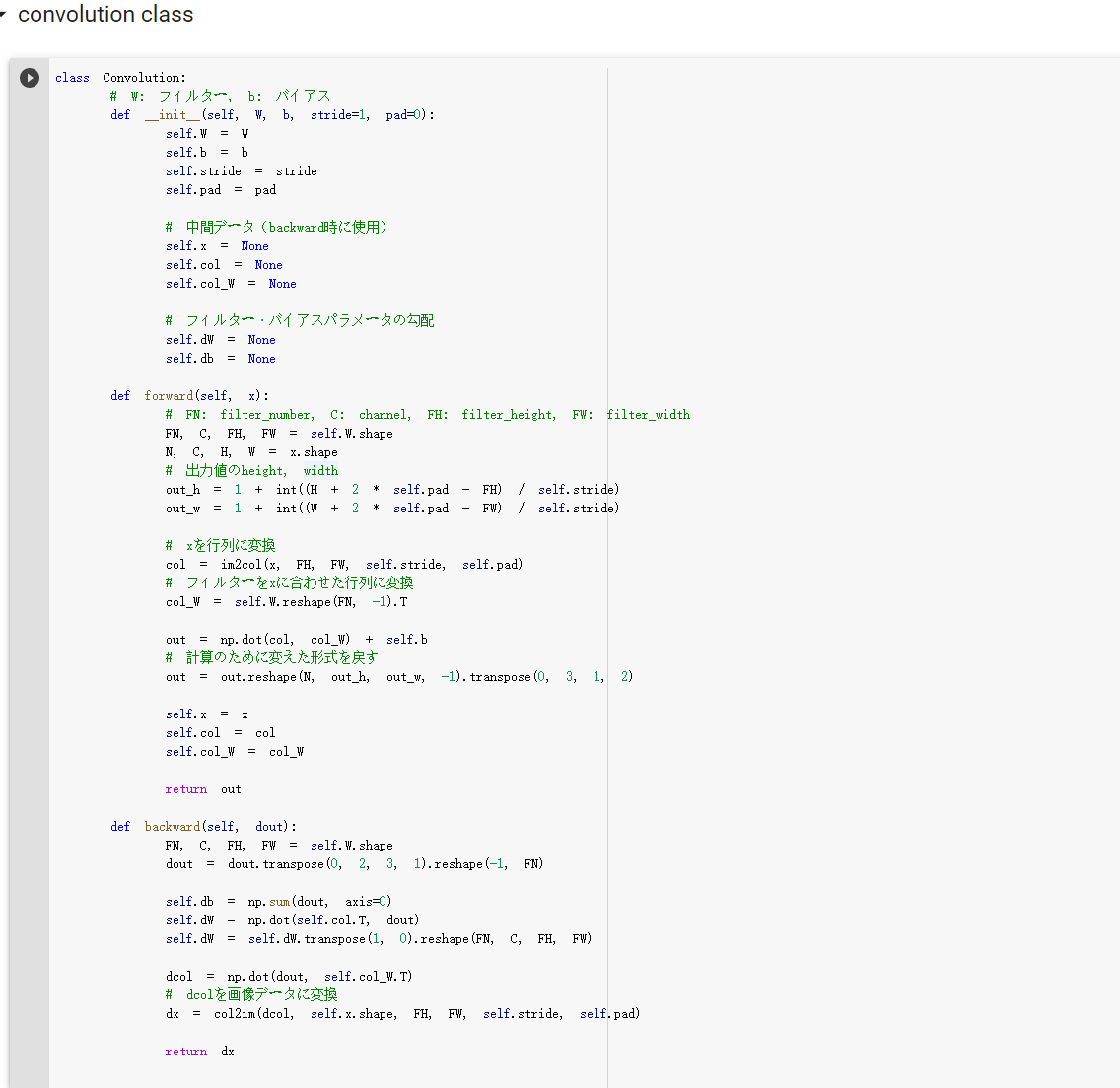
畳み込み層を追加することで、3次元の空間情報も学習できるようになりました。画像の場合、縦、横、チャンネルの3次元のデータをそのまま学習し、次に伝えることができる。数学的計算方式はいつもの結合層と違うところがですが、似てる部分もある。フィルターと畳み込み計算方式を理解することは重要。畳み込み層の後にプーリング層が存在するのは一般的である、違う種類のプーリング層の計算方法を把握する必要がある。

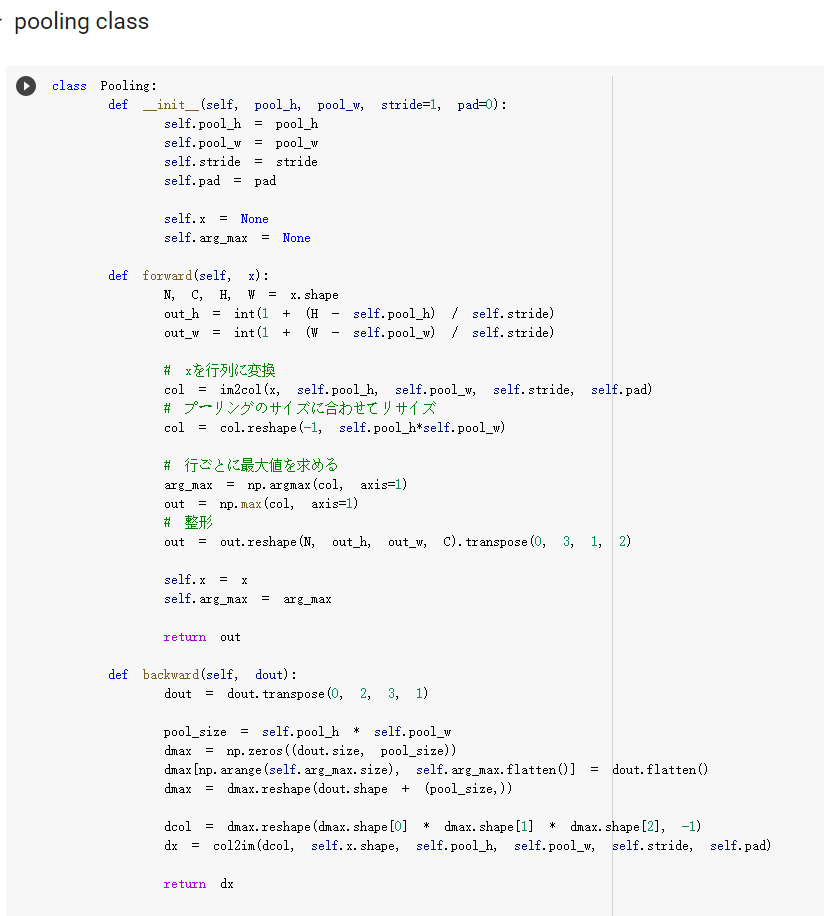
**確認テスト：**

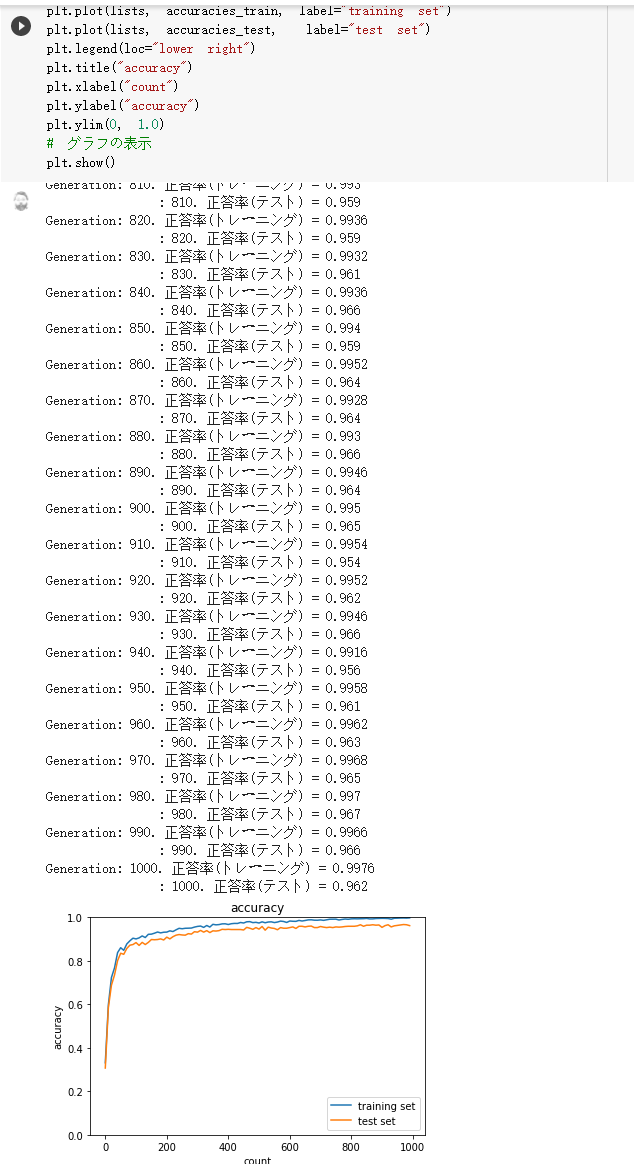


**答え：出力画像のサイズ３ｘ３である。**

**実装演習の確認：**







**まとめ要点５：**

　AlexNetのモデル

AlexNetのモデルの特徴は初の 5 つは畳み込み層で、そのうちのいくつかに max-pooling 層が続き、最後の 3 つは全結合層だった[2]。活性化関数には、非飽和型の ReLU を使用していた。AlexNet は、2012 年 9 月 30 日に開催された ILSVRC 2012[3] に参加した。AlexNet はエラー率 15.3% で優勝し、次点よりも 10.8% 以上低かった。この論文の主な内容は、モデルの深さが高性能には不可欠であるというもので、計算コストは高くなるものの、GPU を用いて学習することで実現した。