確率勾配降下法
Stochastic Gradient Descent

確率勾配降下法：Stochastic Gradient Descent

❑ バッチ学習 vs. ミニバッチ / オンライン学習

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E_n\left(\mathbf{w}^{(\tau)}\right)$$

データサンプル

学習データセット（TDS）

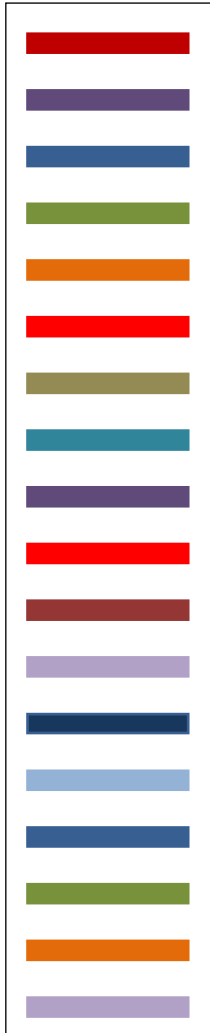# GD: バッチ学習

repeat {

    Evaluate ∇E with all samples

        in TDS (Training dataset);

    Update w;

} until convergence;

GD: バッチ学習

repeat {

Evaluate ∇E with all samples

in TDS (Training dataset);

Update w;

} until convergence;

GD: バッチ学習

repeat {
　　　Evaluate ∇E with all samples
　　　　　　in TDS (Training dataset);
　　　Update w;
} until convergence;

SGD: オンライン学習

```
repeat {
        for each sample in TDS {
                Get one sample from TDS;
                Evaluate ∇E with the sample;
                Update w;
        }
        (shuffle TDS;)
} until convergence;
```

SGD: オンライン学習

```
repeat {
    for each sample in TDS {
        Get one sample from TDS;
        Evaluate ∇E with the sample;
        Update w;
    }
    (shuffle TDS;)
} until convergence;
```

SGD: オンライン学習

repeat {

      for each sample in TDS {

            Get one sample from TDS;

            Evaluate ∇E with the sample;

            Update w;

      }

    （shuffle TDS;）

} until convergence;

SGD: オンライン学習

```
repeat {
        for each sample in TDS {
                Get one sample from TDS;
                Evaluate ∇E with the sample;
                Update w;
        }
```

(shuffle TDS;)

```
} until convergence;
```

SGD: オンライン学習

repeat {

　　　for each sample in TDS {
　　　　　Get one sample from TDS;
　　　　　Evaluate ∇E with the sample;
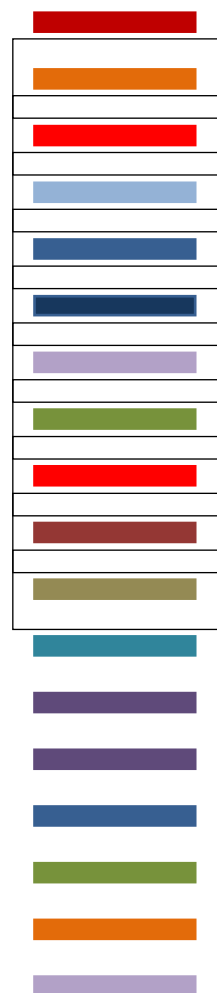　　　　　Update w;
　　　}
　　（shuffle TDS;）
} until convergence;

SGD: オンライン学習

```
repeat {
        for each sample in TDS {
                Get one sample from TDS;
                Evaluate ∇E with the sample;
                Update w;
        }
        (shuffle TDS;)
} until convergence;
```

SGD: ミニバッチ学習

```
repeat {
    for each MB(mini-batch) in TDS {
        Get one MB;
        Evaluate ∇E with the MB;
        Update w;
    }
    (shuffle MB;)
} until convergence;
```

SGD: ミニバッチ学習

repeat {

       for each MB(mini-batch) in TDS {

            Get one MB;

            Evaluate ∇E with the MB;

            Update w;

       }

     (shuffle MB;)

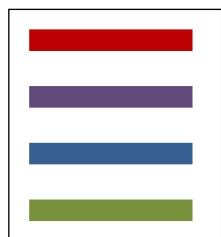} until convergence;

SGD: ミニバッチ学習

repeat {

    for each MB(mini-batch) in TDS {

        Get one MB;

        Evaluate ∇E with the MB;

        Update w;
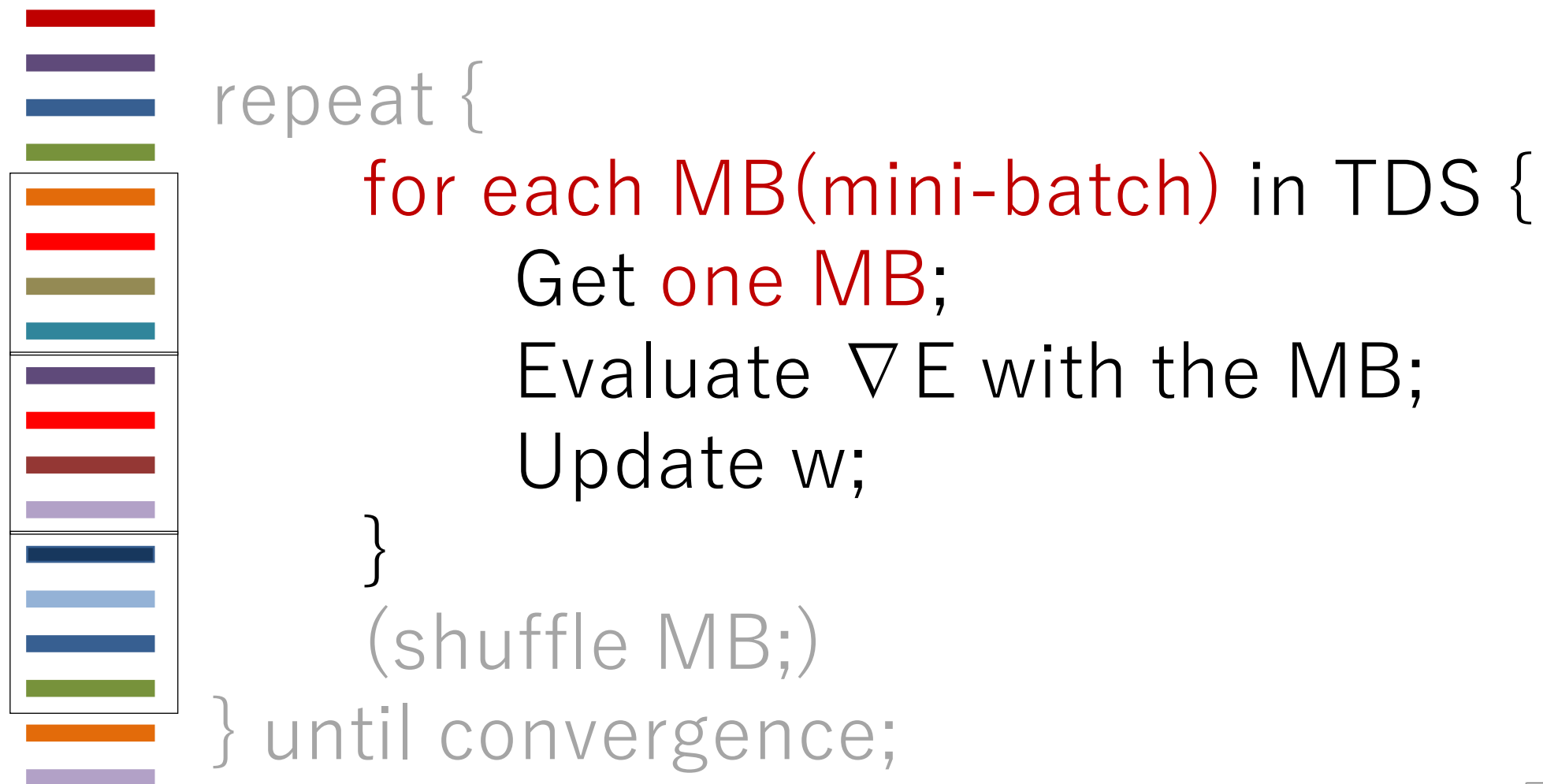
    }

   (shuffle MB;)

} until convergence;

SGD: ミニバッチ学習

repeat {

    for each MB(mini-batch) in TDS {

        Get one MB;

        Evaluate ∇E with the MB;

        Update w;

    }

    (shuffle MB;)

} until convergence;

SGD: ミニバッチ学習

```
repeat {
        for each MB(mini-batch) in TDS {
                Get one MB;
                Evaluate ∇E with the MB;
                Update w;
        }
        (shuffle MB;)
} until convergence;
```

何故 オンライン/ミニバッチか？

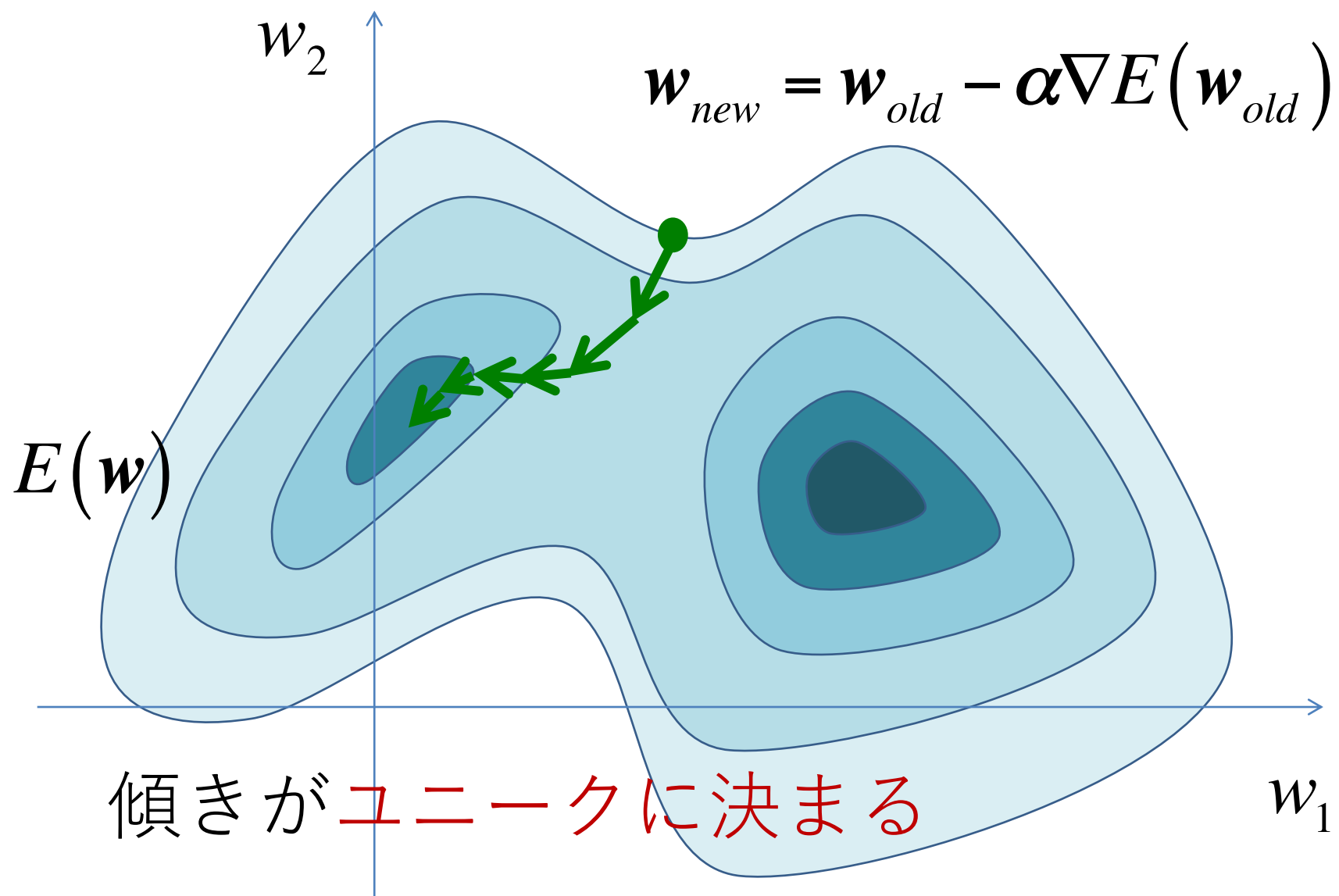オンライン/ミニバッチ(SGD) >> バッチ(GD)

一般にSGDの性能は，GDより数段良い。

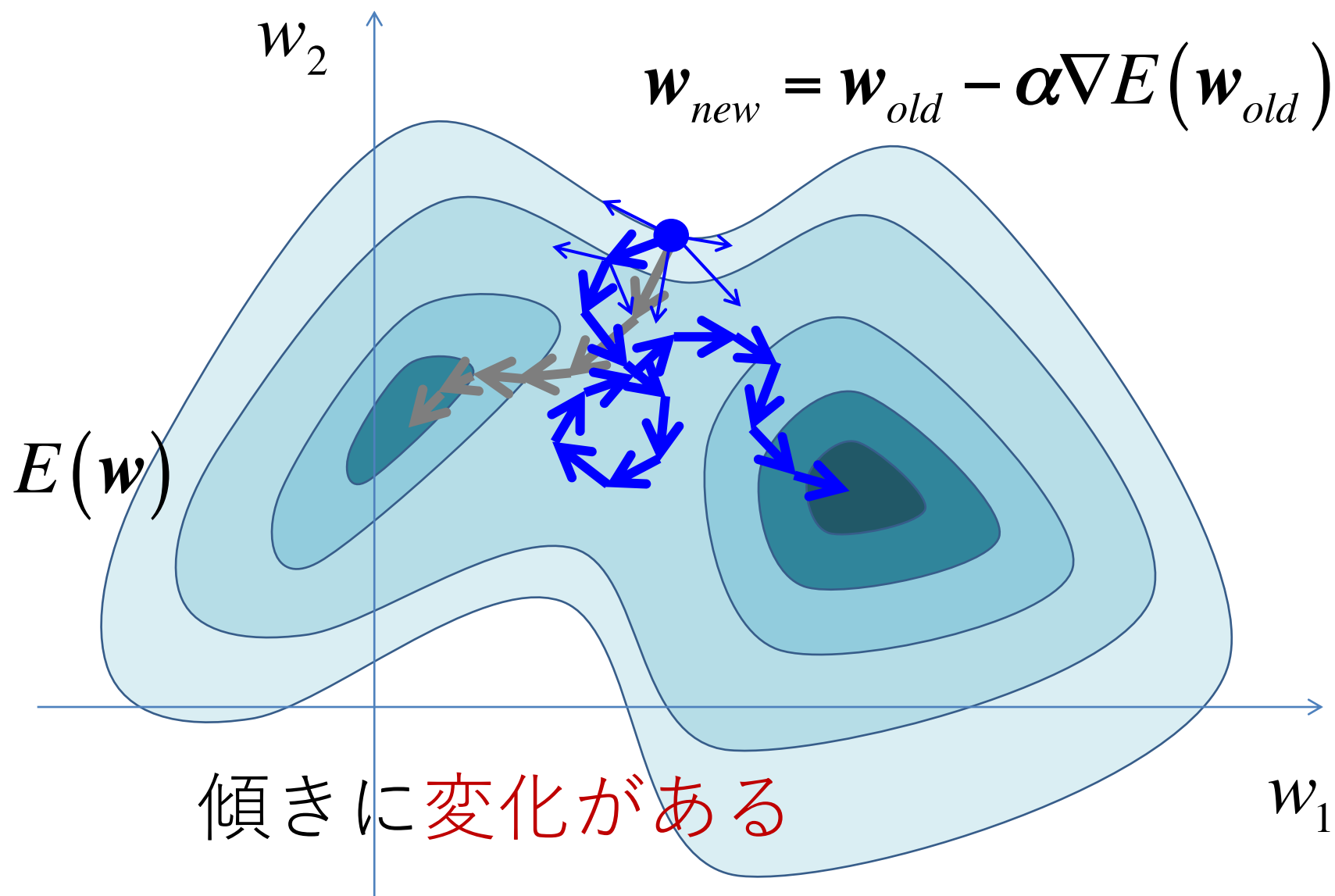バッチ：傾きは，ユニークに決まる

→ 局所最適解にトラップされやすい

オンライン/ミニバッチ: 傾きは，変化に富む

→ 局所最適解にトラップされにくい

最急降下法 (Gradient descent)

$$w_{new} = w_{old} - \alpha \nabla E(w_{old})$$

w_2

$E(w)$

w_1

傾きがユニークに決まる

確率勾配降下法 (Stochastic Gradient Descent)

$$w_{2}$$

$$w_{new} = w_{old} - \alpha \nabla E(w_{old})$$

$$E(w)$$

傾きに変化がある

$$w_{1}$$

まとめ

- 勾配法では，局所最適解にトラップされる可能性を減じることが重要である。
- SGDでは，勾配を求めるために利用するデータを確率的に選ぶことで，局所最適解にトラップされる可能性を減じる。