traingd

Gradient descent backpropagation

Syntax

```
net.trainFcn = 'traingd'
[net,tr] = train(net,...)
```

Description

`traingd` is a network training function that updates weight and bias values according to gradient descent.

`net.trainFcn = 'traingd'` sets the network `trainFcn` property.

`[net,tr] = train(net,...)` trains the network with `traingd`.

Training occurs according to `traingd` training parameters, shown here with their default values:

`net.trainParam.epochs`	`1000`	Maximum number of epochs to train
`net.trainParam.goal`	`0`	Performance goal
`net.trainParam.showCommandLine`	`false`	Generate command-line output
`net.trainParam.showWindow`	`true`	Show training GUI
`net.trainParam.lr`	`0.01`	Learning rate
`net.trainParam.max_fail`	`6`	Maximum validation failures
`net.trainParam.min_grad`	`1e-5`	Minimum performance gradient
`net.trainParam.show`	`25`	Epochs between displays (NaN for no displays)
`net.trainParam.time`	`inf`	Maximum time to train in seconds

Network Use

You can create a standard network that uses `traingd` with `feedforwardnet` or `cascadeforwardnet`. To prepare a custom network to be trained with `traingd`,

1. Set `net.trainFcn` to `'traingd'`. This sets `net.trainParam` to `traingd`'s default parameters.
2. Set `net.trainParam` properties to desired values.

In either case, calling `train` with the resulting network trains the network with `traingd`.

See `help feedforwardnet` and `help cascadeforwardnet` for examples.

More About

collapse all

Gradient Descent Backpropagation

The batch steepest descent training function is `traingd`. The weights and biases are updated in the direction of the negative gradient of the performance function. If you want to train a network using batch steepest descent, you should set the network `trainFcn` to `traingd`, and then call the function `train`. There is only one training function associated with a given network.

There are seven training parameters associated with `traingd`:

- `epochs`
- `show`

- goal
- time
- min_grad
- max_fail
- lr

The learning rate `lr` is multiplied times the negative of the gradient to determine the changes to the weights and biases. The larger the learning rate, the bigger the step. If the learning rate is made too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge. See page 12-8 of [HDB96] for a discussion of the choice of learning rate.

The training status is displayed for every `show` iterations of the algorithm. (If `show` is set to `NaN`, then the training status is never displayed.) The other parameters determine when the training stops. The training stops if the number of iterations exceeds `epochs`, if the performance function drops below `goal`, if the magnitude of the gradient is less than `mingrad`, or if the training time is longer than `time` seconds. `max_fail`, which is associated with the early stopping technique, is discussed in Improving Generalization.

The following code creates a training set of inputs `p` and targets `t`. For batch training, all the input vectors are placed in one matrix.

```
p = [-1 -1 2 2; 0 5 0 5];
t = [-1 -1 1 1];
```

Create the feedforward network.

```
net = feedforwardnet(3,'traingd');
```

In this simple example, turn off a feature that is introduced later.

```
net.divideFcn = '';
```

At this point, you might want to modify some of the default training parameters.

```
net.trainParam.show = 50;
net.trainParam.lr = 0.05;
net.trainParam.epochs = 300;
net.trainParam.goal = 1e-5;
```

If you want to use the default training parameters, the preceding commands are not necessary.

Now you are ready to train the network.

```
[net,tr] = train(net,p,t);
```

The training record `tr` contains information about the progress of training.

Now you can simulate the trained network to obtain its response to the inputs in the training set.

```
a = net(p)
a =
   -1.0026   -0.9962    1.0010    0.9960
```

Try the *Neural Network Design* demonstration `nnd12sd1` [HDB96] for an illustration of the performance of the batch gradient descent algorithm.

Algorithms

`traingd` can train any network as long as its weight, net input, and transfer functions have derivative functions.

Backpropagation is used to calculate derivatives of performance `perf` with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent:

```
dX = lr * dperf/dX
```

Training stops when any of these conditions occurs:

- The maximum number of `epochs` (repetitions) is reached.
- The maximum amount of `time` is exceeded.
- Performance is minimized to the `goal`.
- The performance gradient falls below `min_grad`.
- Validation performance has increased more than `max_fail` times since the last time it decreased (when using validation).

See Also

`traingda` | `traingdm` | `traingdx` | `trainlm`

Introduced before R2006a