Component Derivations



Let the input be x = and the class label or the desired output be “d.”

Let be the local induced vector of Layer-1.

Let be the bias vector of layer 1.

Let be the weight matrix of Layer-1.

Let be the activation output of Layer-1

Let be the local induced vector of Layer-2.

Let be the bias vector of layer 2.

Let be the weight matrix of layer-2.

Let be the actual or activation output of Layer-2.

Let the cost function be (1.0)

Output Layer or Layer-2:

Consider the gradient of the cost function *C* with respect to weight .

(1.1)

(1.2)

(1.3)

(1.4)

Substituting eqns. 1.2, 1.3 and 1.4 in eqn. 1.1, we have the gradient of the cost function *C* with respect to weight as

(1.5)

Observe is the partial derivative of the cost function with respect to input or induced local field of the neuron.

Similarly,

(1.6)

(1.7)

since

**Neuron-1, Hidden Layer -1:**

(2.1)

(2.2)

(2.3)

(2.4)

(2.5)

Substituting eqns. 2.4 and 2.5 in eqns. 2.3, we have

(2.6)

Substituting eqns. 1.2, 1.3, 2.2 and 2.6 in eqn. 2.1, we have

(2.7)

Observe is the partial derivative of the cost function with respect to input or induced local field of the neuron in the hidden layer.

Similarly,

(2.8)

(2.9)

Substituting eqns. 1.2, 1.3, 2.2 and 2.9 in eqn. 2.8, we have

(2.10)

Bias Update:

(2.11)

Substituting eqns. 1.2, 1.3, and 2.2 in eqn. 2.11, we have

(2.12)

(2.13)

Substituting eqn. 2.13 in eqn. 2.12, we have

(2.14)

**For Neuron-2, Hidden layer-1, we have**

(2.15)

(2.16)

(2.17)

Observe the partial derivative of the cost function w.r.t a weight is the product of the partial derivative of the cost function w.r.t to the local induced field and the input through that weight.

**Matrix-Vector Form Derivations:**

The local gradient vector of layer-2 from the partial derivative of the cost function with respect to input is

1.1

The local gradient vector of layer-1 from the partial derivative of the cost function with respect to input is

1.2

Weight Updates:

where includes the weight of the bias and includes the bias input.

1.3

1.4

1.5

1.6



**Matrix-Vector Form Derivations:**

The local gradient vector of layer-2 from the partial derivative of the cost function with respect to input is

2.1

2.2

Weight Updates:

where includes the weight of the bias and includes the bias input.

2.3

2.4

2.5

2.6

Let there be layers in the network with *L* being the output layer.

Let in each layer there be *K* neurons.

Let each layer weight matrices without the bias be

The local gradient vector of layer-*L* from the partial derivative of the cost function with respect to input is