**Gradient Descent in Linear Models & Loss Function**

**1. Why This Matters**

To create working machine learning models that can adapt to different problems, you need to understand the **key drivers** of how algorithms learn.  
This lecture builds on what you already know: linear models, loss functions, and gradient descent.

**2. The Linear Model Recap**

* Output yi = input xi × weight w + bias b.
* Each output yi corresponds to a single observation (like one apartment’s price).
* The inputs, weights, and bias stay consistent for all observations; only xi​ and yi​ change with each observation.
* We compare yi​ to the **target** ti (the true value we want to predict).

**3. Choosing the Loss Function**

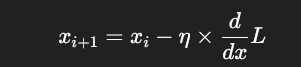
* The loss function, denoted L(y,t) measures how far outputs are from targets.
* Different notations exist: LLL for loss, CCC for cost, EEE for error — they mean the same thing.
* For regression problems, the **L2-norm loss** (squared error) is common:



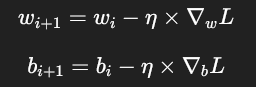
Dividing by 2 is a mathematical trick to simplify later derivative calculations — it doesn’t affect the behavior.

**4. Moving from One Dimension to Many Dimensions**

* Previously, gradient descent worked on a single variable xxx:



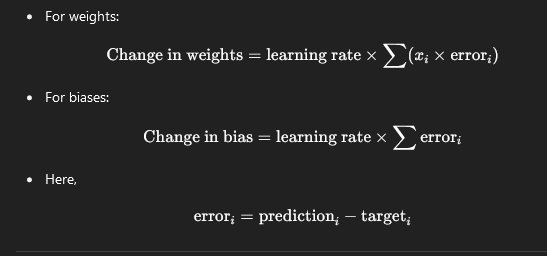
For machine learning models, weights www and biases bbb are **vectors or matrices**, so the update rule becomes:



Here, ∇wL means the **gradient of the loss with respect to weights**, and similarly for bias.

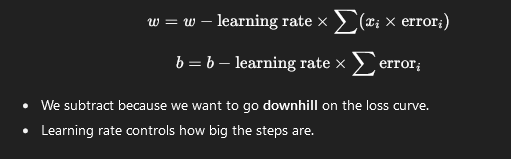
**5. Calculating the Gradient for Weights**

* The gradient of the loss function w.r.t. weights www is:



**6. Update rule — how to adjust weights and biases**

Each time we update:



**7. Repeat and stop**

* We repeat these updates many times until the loss stops changing much.
* Then the model is trained and ready to make good predictions.

**In short:**

* Use inputs, weights, and bias to make predictions.
* Check how wrong predictions are with a loss function.
* Use gradient descent to change weights and bias bit by bit to reduce loss.
* Repeat until the loss is as small as possible.