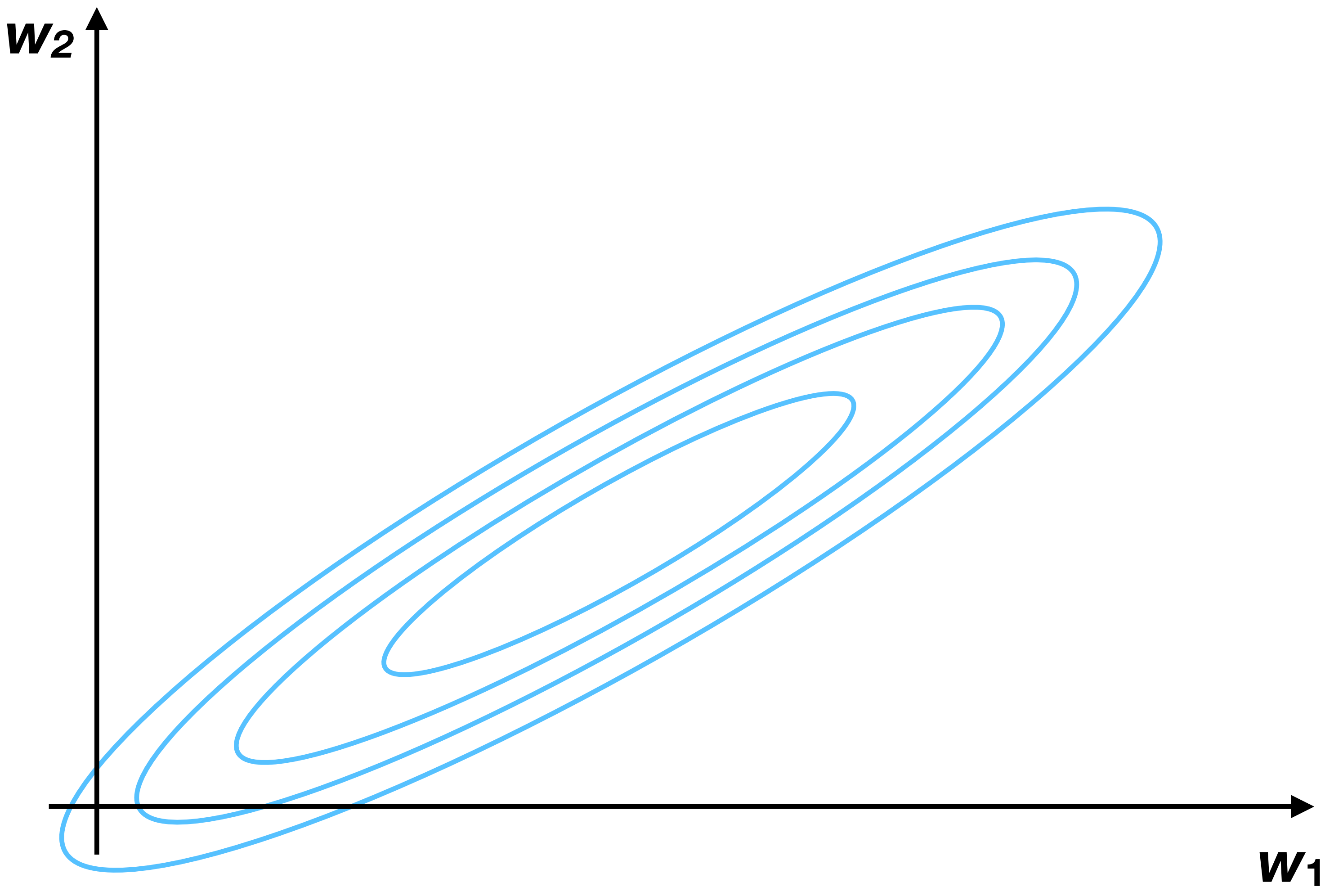
CS 4342: Class 10

Jacob Whitehill

Curvature of the objective function

Curvature of the objective function

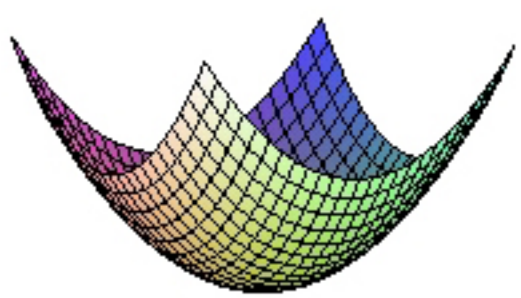


Convex ML models

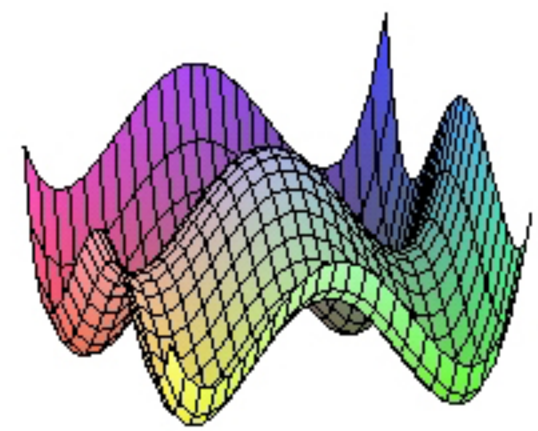
Convex ML models

- The two main ML models we have examined — linear regression and softmax regression — have loss functions that are **convex**.

- With a convex function *f*, every local minimum is also a global minimum.



convex



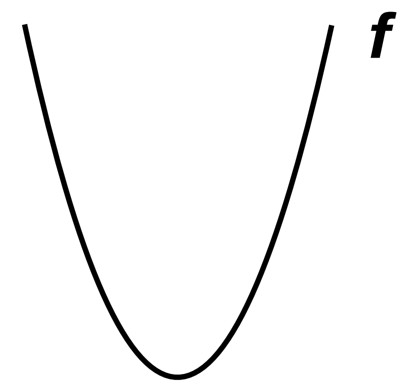
non-convex

- Convex functions are ideal for conducting gradient descent.

https://plus.maths.org/content/convexity

Convexity in 1-d

- How can we tell if a 1-d function *f* is convex?

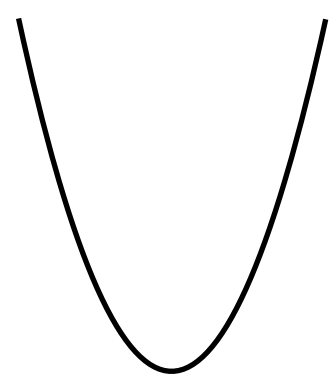


- What property of *f* ensures there is only one local minimum?

Jacob Whitehill, WPI

Convexity in 1-d

- How can we tell if a 1-d function *f* is convex?

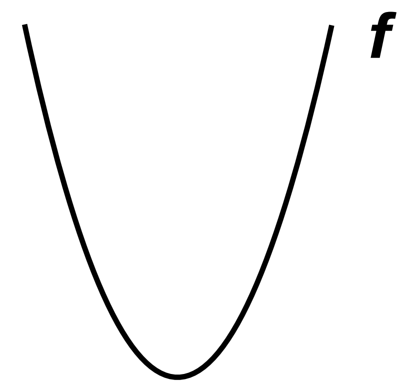


- What property of *f* ensures there is only one local minimum?

    - From left to right, the slope of *f never decreases*.

Convexity in 1-d

- How can we tell if a 1-d function *f* is convex?

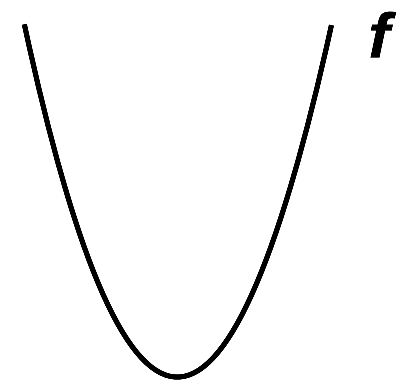


- What property of *f* ensures there is only one local minimum?

    - From left to right, the slope of *f* *never decreases*.
    ==> the derivative of the slope is always non-negative.

Convexity in 1-d

- How can we tell if a 1-d function *f* is convex?



- What property of *f* ensures there is only one local minimum?

    - From left to right, the slope of *f* *never decreases*.
      ==> the derivative of the slope is always non-negative.
      ==> the second derivative of *f* is always non-negative.

Convexity in higher dimensions

- For higher-dimensional *f*, convexity is determined by the second derivative matrix, known as the **Hessian** of *f*.

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

- For $f : \mathbb{R}^m \rightarrow \mathbb{R}$, *f* is convex if the Hessian matrix is positive semi-definite (PSD) for *every* input **x.**

Positive semi-definite

- Positive semi-definite is the matrix analog of being “non-negative”.

- A real symmetric matrix **A** is **positive semi-definite (PSD)** if (equivalent conditions):

Positive semi-definite

- Positive semi-definite is the matrix analog of being “non-negative”.

- A real symmetric matrix **A** is **positive semi-definite (PSD)** if (equivalent conditions):
    - For every vector **v**: $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$
        - Therefore: If there exists *any* vector **v** such that $\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$, then **A** is *not* PSD.

Positive semi-definite

- Positive semi-definite is the matrix analog of being “non-negative”.

- A real symmetric matrix **A** is **positive semi-definite (PSD)** if (equivalent conditions):
  - For every vector **v**: $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$
    - Therefore: If there exists *any* vector **v** such that $\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$, then **A** is *not* PSD.
  - All its eigenvalues are ≥ 0.
    - If **A** happens to be diagonal, then its eigenvalues are the diagonal elements.

Example

- Suppose $f(x, y) = 3x^2 + 2y^2 - 2$.
- Then the first derivatives are: $\frac{\partial f}{\partial x} = 6x \quad \frac{\partial f}{\partial y} = 4y$

Example

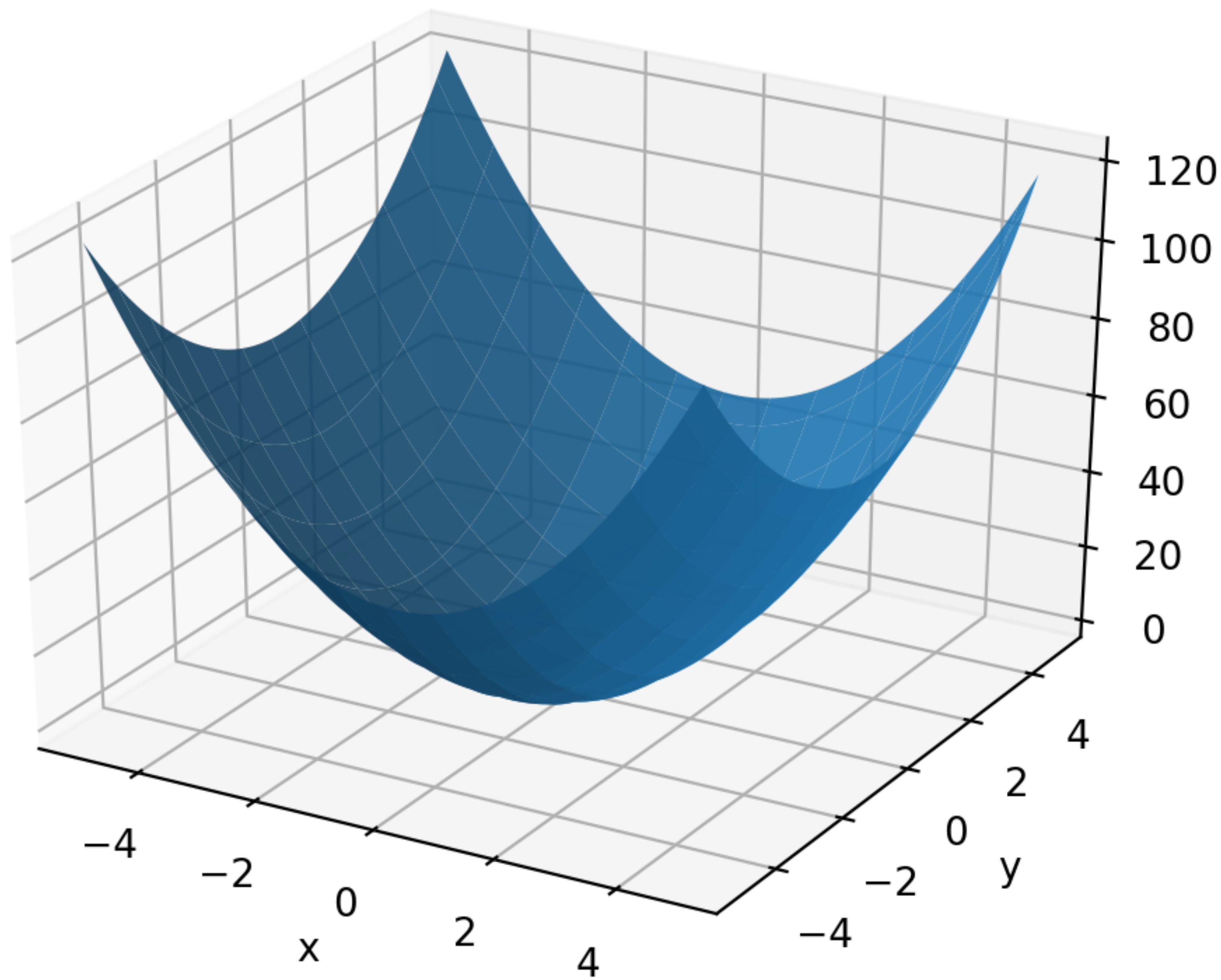
- Suppose $f(x, y) = 3x^2 + 2y^2 - 2$.
- Then the first derivatives are: $\frac{\partial f}{\partial x} = 6x \quad \frac{\partial f}{\partial y} = 4y$
- The Hessian matrix is therefore:

$$\mathbf{H} = \left[\begin{array}{cc} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial y} \end{array} \right] = \left[\begin{array}{cc} 6 & 0 \\ 0 & 4 \end{array} \right]$$

- Notice that **H** for this *f* does not depend on (x,y).
- Also, **H** is a diagonal matrix (with 6 and 4 on the diagonal). Hence, the eigenvalues are just 6 and 4. Since they are both non-negative, then *f* is convex.

Example

- Graph of $f(x, y) = 3x^2 + 2y^2 - 2$:



Example

- Recall: if $\mathbf{H}(x, y)$ is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^T\mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* $\mathbf{v}$
  - All eigenvalues of $\mathbf{H}(x,y)$ are non-negative.

- Which of the following function(s) are convex?
  - $x^2 + y + 5$
  - $x^2 + 3xy$
  - $x^4 + xy + x^2$

Example

- Recall: if **H**(x, y) is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
    - $\mathbf{v}^T\mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* **v**
    - All eigenvalues of **H**(x,y) are non-negative.
- Which of the following function(s) are convex?
    - $x^2 + y + 5$ $\quad \mathbf{H} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$
    - $x^2 + 3xy$
    - $x^4 + xy + x^2$

Example

- Recall: if $\mathbf{H}(x, y)$ is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^T\mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* $\mathbf{v}$
  - All eigenvalues of $\mathbf{H}(x,y)$ are non-negative.

- Which of the following function(s) are convex?
  - $x^2 + y + 5$ $\quad \mathbf{H} = \left[\begin{array}{cc} 2 & 0 \\ 0 & 0 \end{array} \right]$ **Eigenvalues are 2, 0 => PSD.**
  - $x^2 + 3xy$
  - $x^4 + xy + x^2$

Example

- Recall: if $\mathbf{H}(x, y)$ is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^T\mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* $\mathbf{v}$
  - All eigenvalues of $\mathbf{H}(x,y)$ are non-negative.

- Which of the following function(s) are convex?
  - $x^2 + y + 5$
  - $x^2 + 3xy$ $\quad \mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix}$
  - $x^4 + xy + x^2$

Example

- Recall: if $\mathbf{H}(x, y)$ is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^\top \mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* $\mathbf{v}$
  - All eigenvalues of $\mathbf{H}(x,y)$ are non-negative.

- Which of the following function(s) are convex?
  - $x^2 + y + 5$
  - $x^2 + 3xy$ $\quad \mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \mathbf{v}^\top \mathbf{H} \mathbf{v} = -4$

    Not PSD.
  - $x^4 + xy + x^2$

Jacob Whitehill, WPI

Example

- Recall: if **H**(x, y) is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^T\mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* **v**
  - All eigenvalues of **H**(x,y) are non-negative.

- Which of the following function(s) are convex?

  - $x^2 + y + 5$

  - $x^2 + 3xy$

  - $x^4 + xy + x^2$ $\quad \mathbf{H} = \begin{bmatrix} 12x^2 + 2 & 1 \\ 1 & 0 \end{bmatrix}$

Jacob Whitehill, WPI

Example

- Recall: if $\mathbf{H}(x, y)$ is the Hessian of f, then f is convex if — at *every* (x,y), we can show (equivalently):
  - $\mathbf{v}^\top \mathbf{H}(x,y)\mathbf{v} \geq 0$ for *every* $\mathbf{v}$
  - All eigenvalues of $\mathbf{H}(x,y)$ are non-negative.

- Which of the following function(s) are convex?
  - $x^2 + y + 5$
  - $x^2 + 3xy$
  - $x^4 + xy + x^2$

$x = 1$

$$\mathbf{H} = \left[\begin{array}{cc} 12x^2+2 & 1 \\ 1 & 0 \end{array}\right] \quad \mathbf{v} = \left[\begin{array}{c} -1 \\ 15 \end{array}\right] \quad \mathbf{v}^\top \mathbf{H}\mathbf{v} = -16$$

Not PSD.

Jacob Whitehill, WPI

Convexity of linear regression and softmax regression

- Why are they convex?

- First, recall that, for any matrices **A**, **B** that can be multiplied:

    - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Convexity of linear regression and softmax regression

- Why are they convex?

- Next, recall the gradient of f_{MSE} (for linear regression):

$$
\begin{aligned}
\nabla_{\mathbf{w}} f_{\mathrm{MSE}} &= \mathbf{X}(\hat{\mathbf{y}} - \mathbf{y}) \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{w} - \mathbf{y}) \\
\mathbf{H} &= \mathbf{X}\mathbf{X}^\top
\end{aligned}
$$

Convexity of linear regression and softmax regression

- Why are they convex?

- Next, recall the gradient of f_{MSE} (for linear regression):

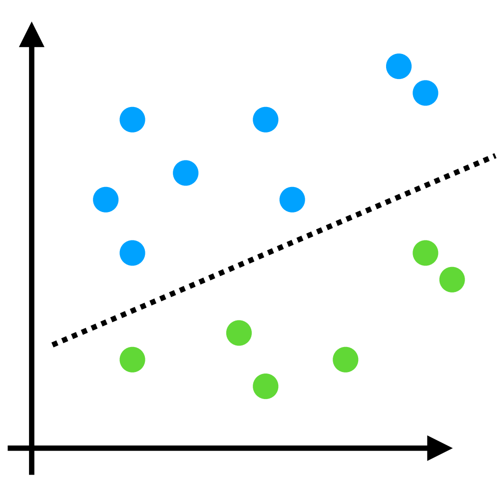
$$
\begin{aligned}
\nabla_{\mathbf{w}} f_{\text{MSE}} &= \mathbf{X}(\hat{\mathbf{y}} - \mathbf{y}) \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{w} - \mathbf{y}) \\
\mathbf{H} &= \mathbf{X}\mathbf{X}^\top
\end{aligned}
$$

- For any vector **v**, we have:

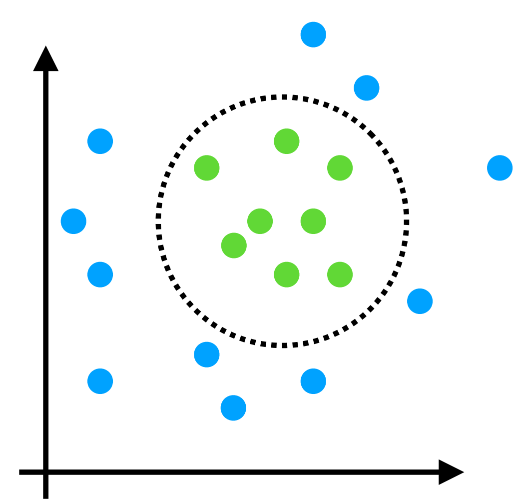
$$
\begin{aligned}
\mathbf{v}^\top \mathbf{X}\mathbf{X}^\top \mathbf{v} &= \left(\mathbf{X}^\top \mathbf{v}\right)^\top \left(\mathbf{X}^\top \mathbf{v}\right) \\
&\geq 0
\end{aligned}
$$

Convex ML models

- Beyond linear regression and softmax regression, what other convex ML models are there?
- One of the most prominent is the **support vector machine (SVM)**.
- SVMs provide a way to classify examples using both linear and non-linear decision boundaries:



Linear decision boundary



Non-linear decision boundary