Lecture 15: Gradient descent – visualization

Ciaran Evans

Logistics

- HW 4 due Monday, February 24
- HW 5 is deliberately short, and several of the questions are optional
- Project 1 is due Friday, March 7

Last time

Gradient descent: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$

- Why the gradient?

Last time

Gradient descent: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$

- Why the gradient? $-\nabla f(\mathbf{x})$ is the **direction of steepest decrease** in f

- How do we choose step size α_k?

Last time

Gradient descent: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$

- Why the gradient? $-\nabla f(\mathbf{x})$ is the **direction of steepest decrease** in f

- How do we choose step size α_k? **maximize benefit**
  - Would like α_k to minimize $f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))$
  - Exact minimization is expensive and unnecessary; instead, find "good enough" α_k with backtracking line search (sufficient decrease condition)

Today: behavior of gradient descent

Motivating example: Data on med school admissions for 55 students

- `GPA`: student's undergraduate GPA
- `MCAT`: student's MCAT score

Function to minimize:

$$f(\beta_0, \beta_1) = \sum_{i=1}^{n} (\mathsf{MCAT}_i - \beta_0 - \beta_1 \mathsf{GPA}_i)^2$$

- Compass search beginning at (0, 0): 5282 iterations
- Gradient descent with backtracking ~~linear~~ line search beginning at (0, 0): 6517 iterations

Motivating question: Why so many iterations?

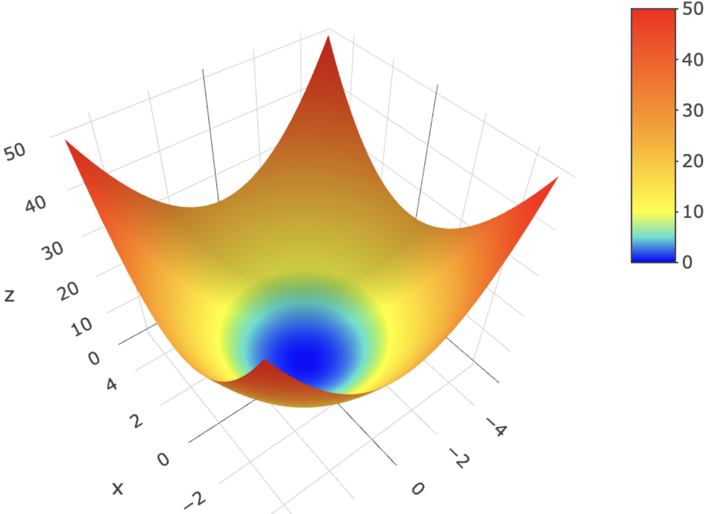
Preliminaries: plotting a function of two variables

$$f(x, y) = x^2 + y^2$$

- Two inputs: x and y
- One output value for each (x, y): $z = f(x, y)$
- In total: function described by 3 coordinates, (x, y, z)

Preliminaries: plotting a function of two variables

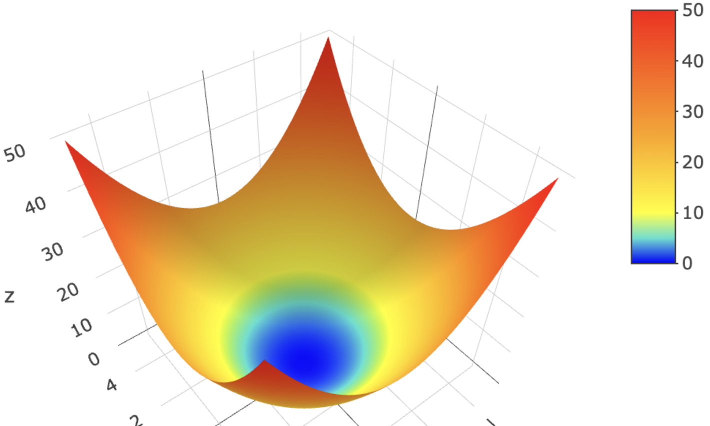
$$f(x, y) = x^2 + y^2$$



Preliminaries: plotting a function of two variables

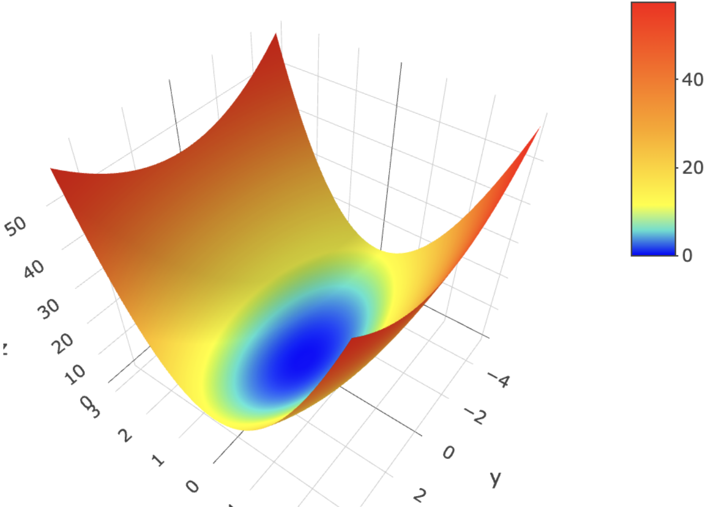
$$f(x, y) = x^2 + y^2$$

Question: What will happen to the shape if we instead plot $5x^2 + 0.5y^2$?



Preliminaries: plotting a function of two variables

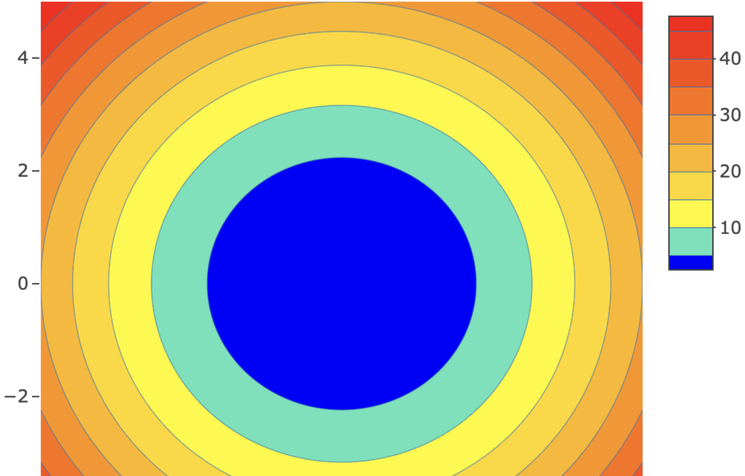
$$f(x, y) = 5x^2 + 0.5y^2$$



Preliminaries: contour plots

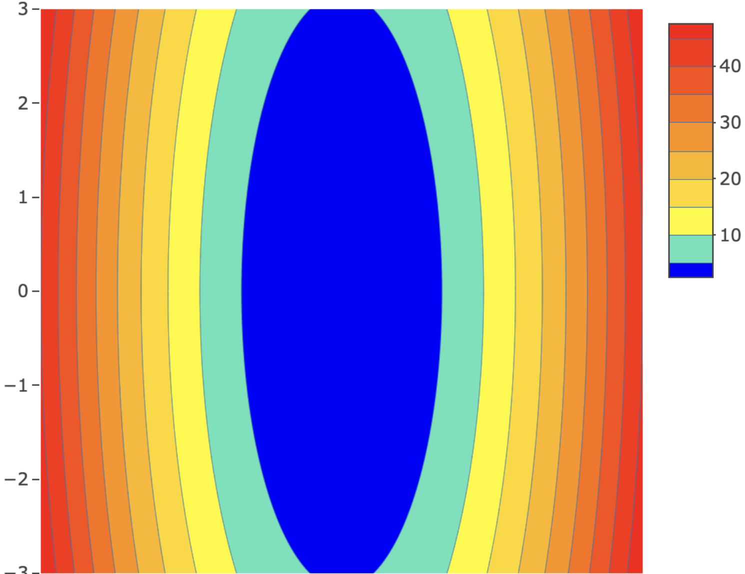
Definition: The *contour* of $f(x, y)$ at c is the set of all points (x, y) for which $f(x, y) = c$

Example: $f(x, y) = x^2 + y^2$



Preliminaries: contour plots

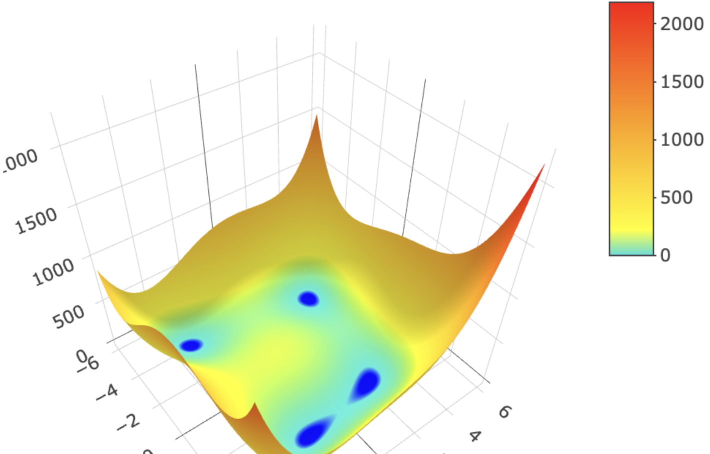
Example: $f(x, y) = 5x^2 + 0.5y^2$



Another example

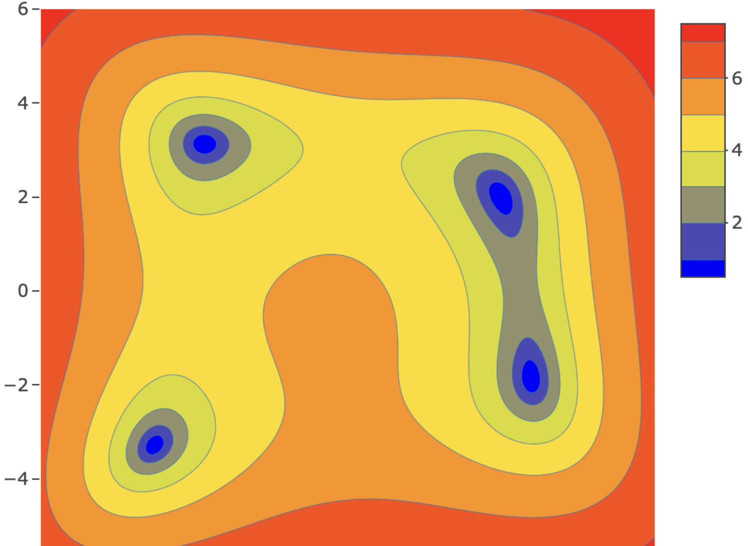
$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

Question: What do you notice about this function?



Another example

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$



Activity

Activity on the course website:

https://sta379-s25.github.io/practice_questions/pq_15.html

- Visualize shape of loss function for regression problem
- Explore how shape of loss function impacts performance of gradient descent