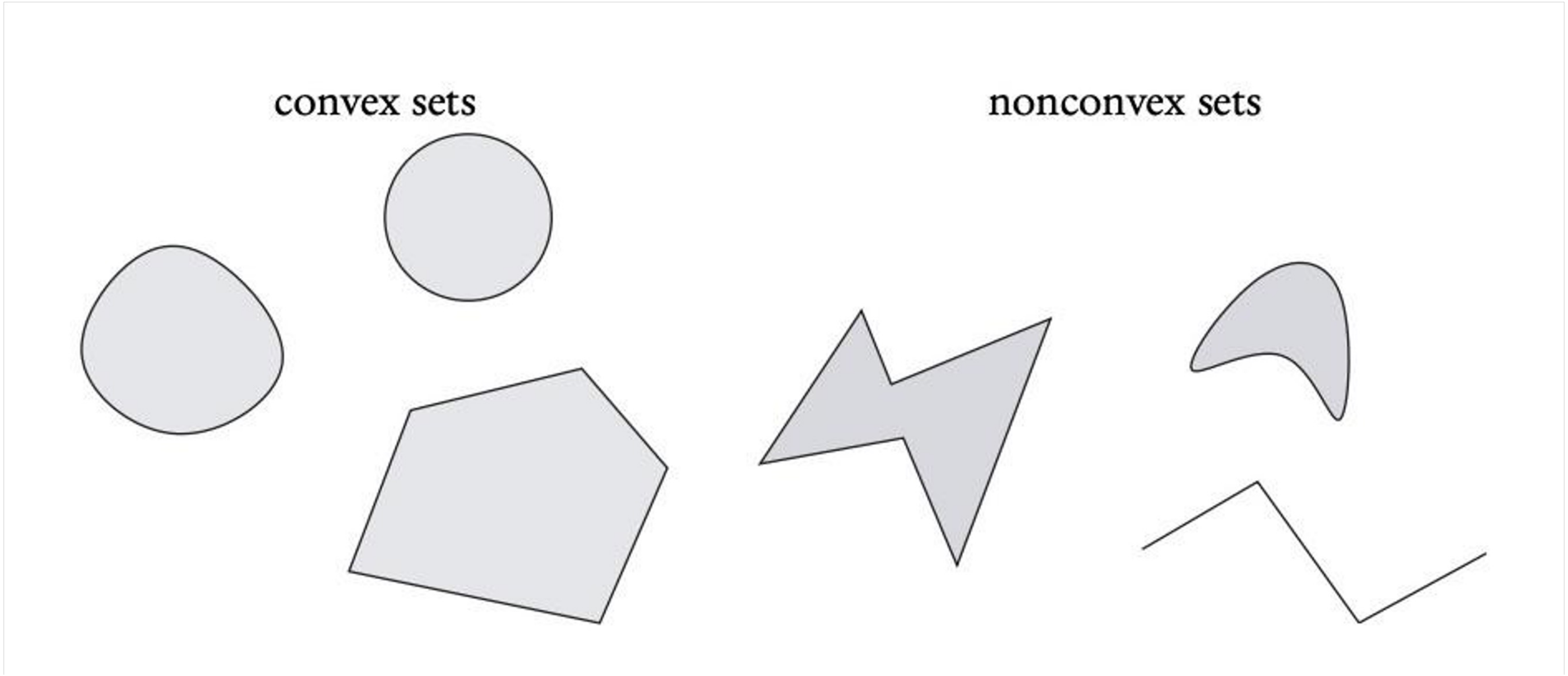
Support Vector Machine

Merisa Kraja and Amanda Olson
Research Mentor: Dr. Tran

Background

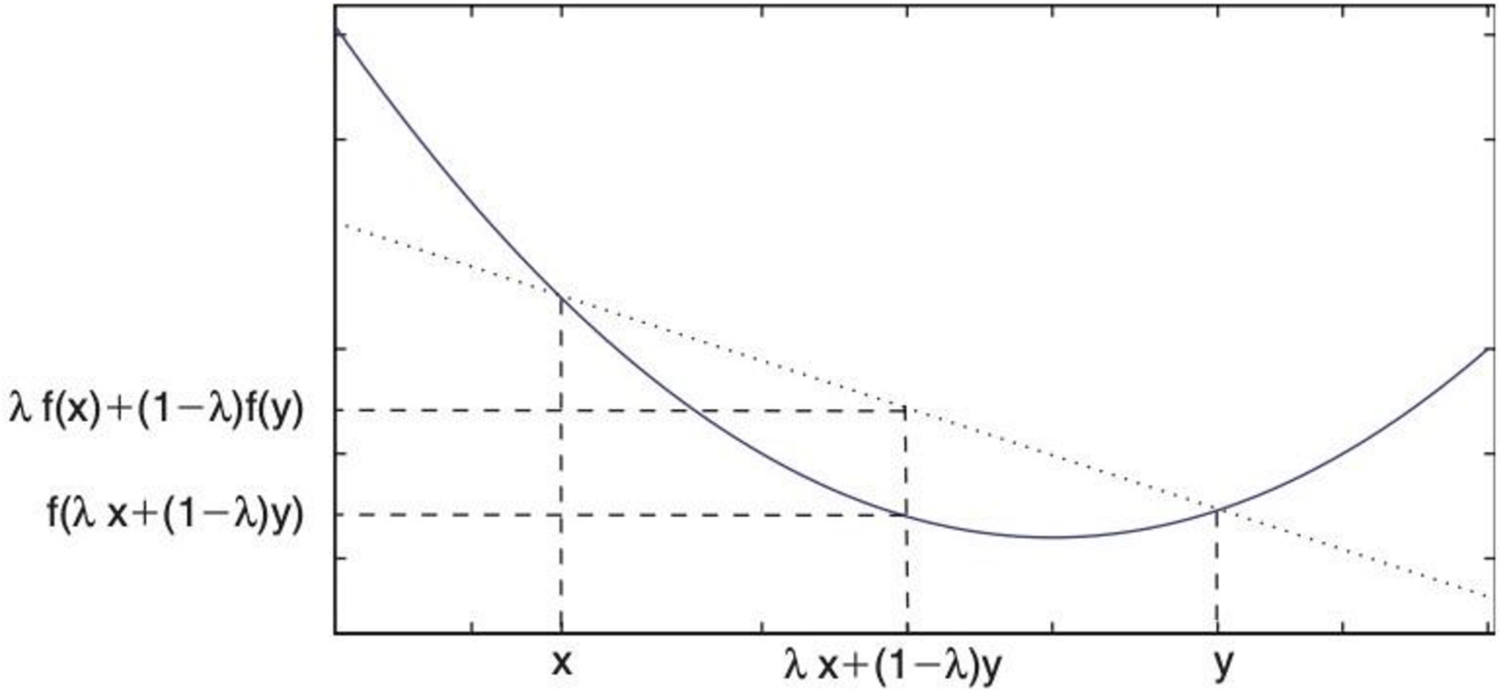
- First Order Optimality Condition: Gradient vanishes at optimal points.
- Gradient Descent is commonly used in optimization and machine learning.
- Optimization methods:
  - gradient descent with different step sizes, stochastic gradient descent, subgradients, and support vector machines.
- Applications: handwriting recognition, email classification, gene classification, computer vision, natural language processing, and bioinformatics.

Convex Set



Convex Function

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$



Schematic Gradient Descent

Consider the unconstrained minimization problem: $\boldsymbol{min\{f(x) : x \in R^n\}}$

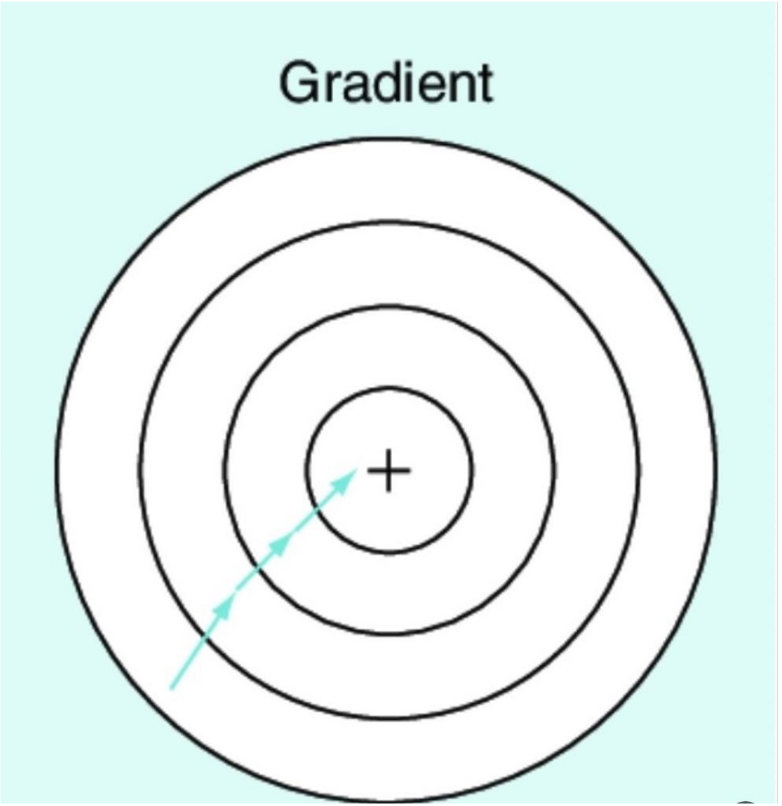
Pick $x0$

Pick $d \in R^n$ where $f'(x;d) = \nabla f(x)^T d < 0$.

Find t such that $f(x+td) < f(x)$

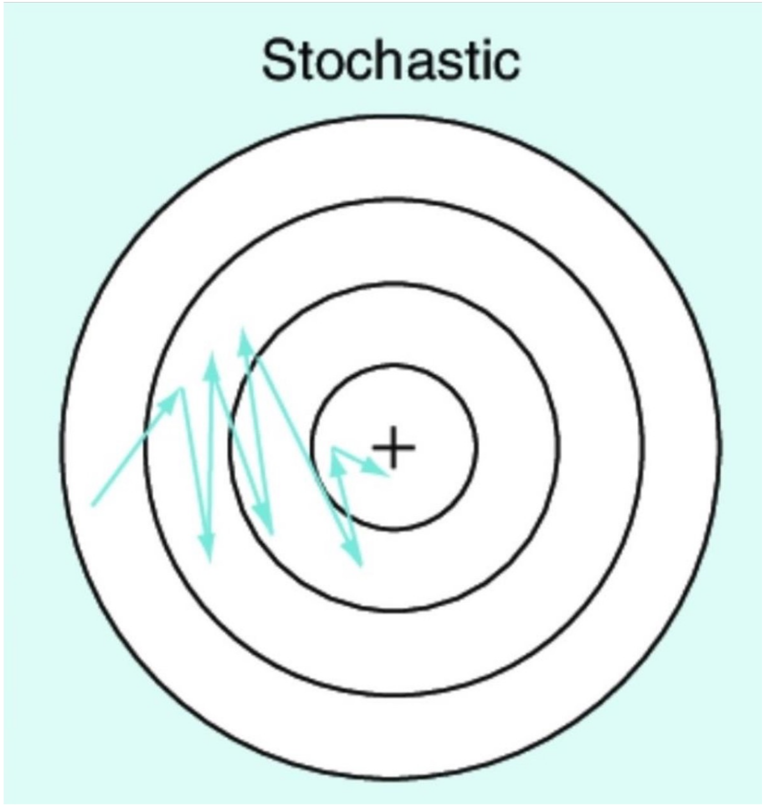
Set $x=x+td$

Continue to descend until tolerance parameter epsilon is reached



Stochastic Gradient Descent

- Chooses a random point or a subset of points from the data
- Reduces computational costs and is useful for larger datasets
- Can lead to poor convergence

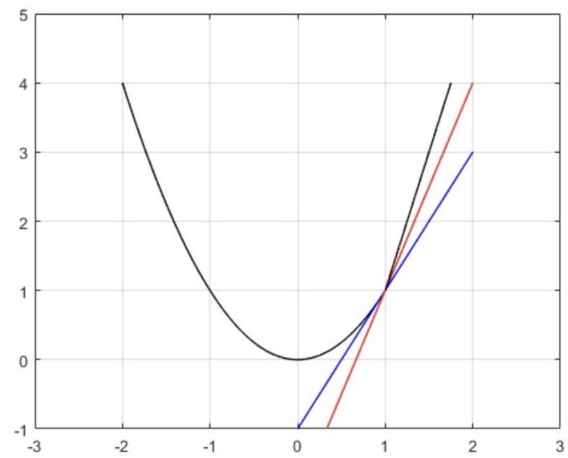


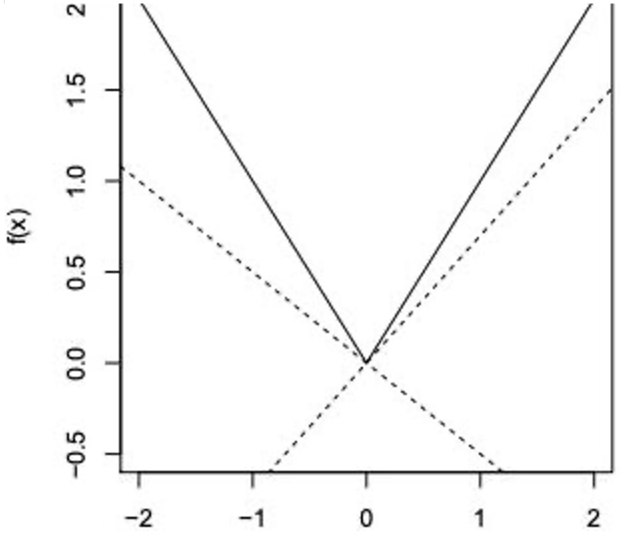
Subgradient

Let $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $x \in dom\ f$. An element $v \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if:

$$< v, x - \bar{x} > \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n$$

The collection of all the subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x}) - \text{subdifferential}$.





Support Vector Machine

Support vector machine is used for classification and regression problems that finds a hyperplane that separates data.

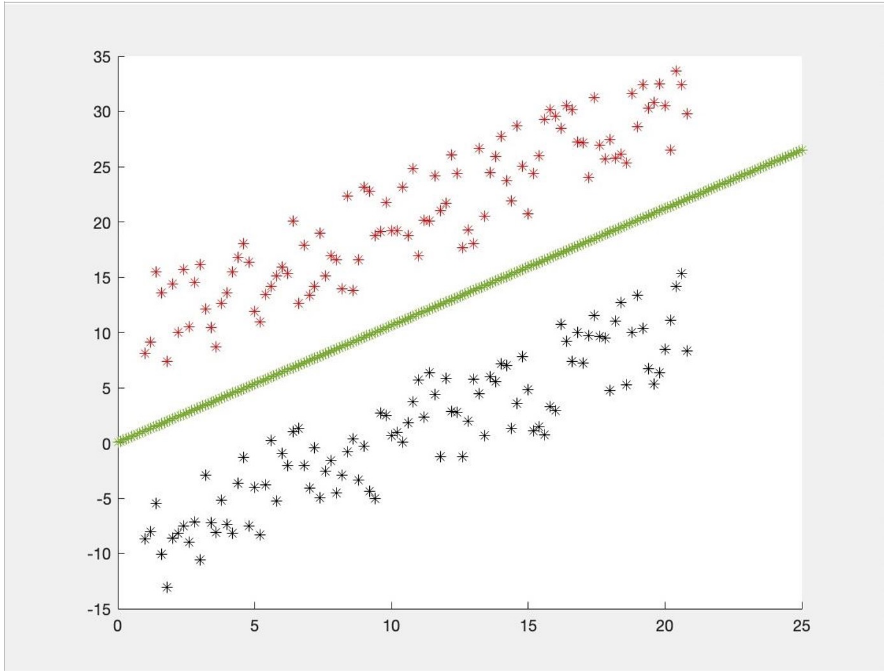
Given S: = $\{(X_i, y_i)\}_{i=1}^{m}$

Hyperplane: *$w^T x + b = 0$*

Optimization problem:

$$minimize\ g(w) \coloneqq \frac{1}{2}\|w\|^2 + \frac{C}{m}\sum_{i=1}^{m} l_i\,(w),\ w \in \mathbb{R}^p$$

where $l_i = max\{0, 1 - y_i X_i w\}$, $C > 0$.



```
% define number of iterations and penalty paramaters
N = 1000;
C =  8.^(-10:10);

% train for each C value specified
for l=1:length(C)
% stochastic subgradient method iterations
for k = 1:N
    %get random index from 1 to m
    temp = randperm(m);
    i = temp(1);
    if data_y(i)*data(i,:)*w'<1
        sub = -C(1)*data_y(i)*data(i,:)+w;
    else
        sub = w;
    end
end
    w = w-1/(k)*sub;
```

Spam Dataset

Attributes:

- Artificial dataset
- 3681 observations for training and 526 observations for testing, with 56 columns each
- Column of 1 and -1 for response variable y

Use:

- Use training set to find hyperplane
- Use test set to assess accuracy of **stochastic subgradient method**
- Analyze penalty parameters/misclassifications

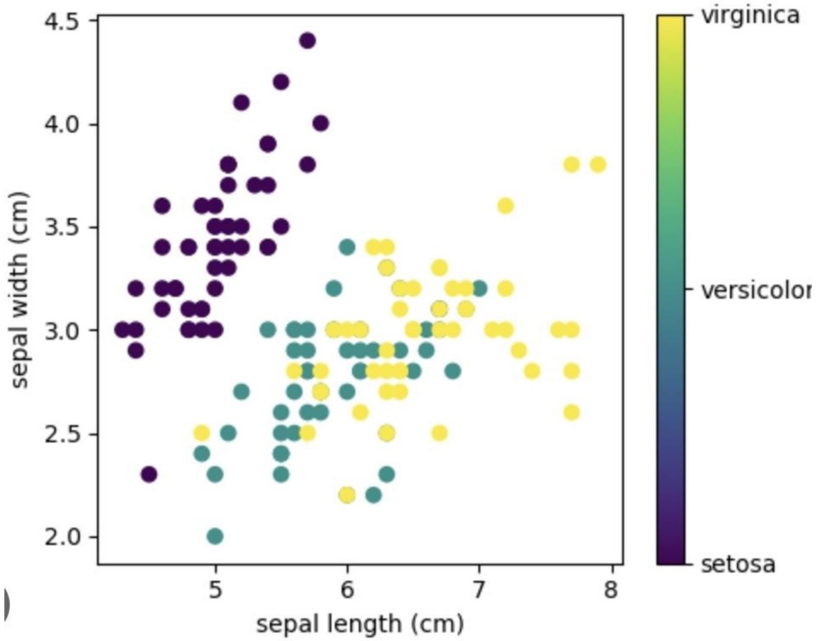
Iris Dataset

Attributes:

- Real dataset
- Has data on sepal length and width, and petal length and width

Use:

- Create training set to find hyperplane between setosa and versicolor
- Create test set to test accuracy of those predictions
- Analyze penalty parameters/misclassifications



Next Steps and Applications

Applications:

- Handwriting recognition
- Gene classification
- Natural language processing
- Computer vision
- Bioinformatics

Next Steps:

- Find better ways to divide the training and testing set data and to pick the soft margins
- Explore clustering and radial basis function kernel

References

Beck, A. *Introduction to Nonlinear Optimization*.

Boyd, S., & Vandenberghe, L. (n.d.). *Convex Optimization*. Cambridge University Press.

Mordukhovich, Boris and Nguyen Mau Nam. *An Easy Path to Convex Analysis and Applications*

Stanford University. *Optimization: Stochastic Gradient Descent*. Unsupervised feature learning and Deep Learning Tutorial. http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/

Support Vector Machine in machine learning. GeeksforGeeks. (2023, May 7). https://www.geeksforgeeks.org/support-vector-machine-in-machine-learning/

Thank you!

Dr. Putonti and CAS-USRE

Dr. Tran

Amanda Olson

Audience