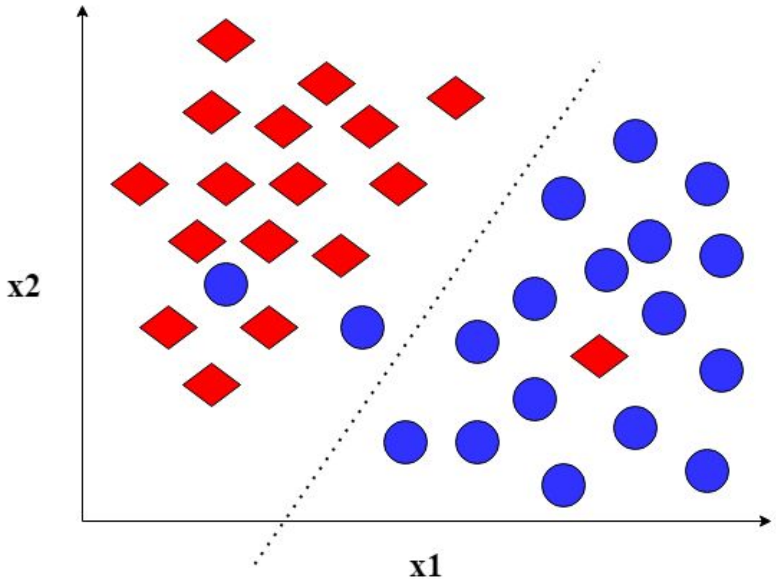
Classification

Binary Classification



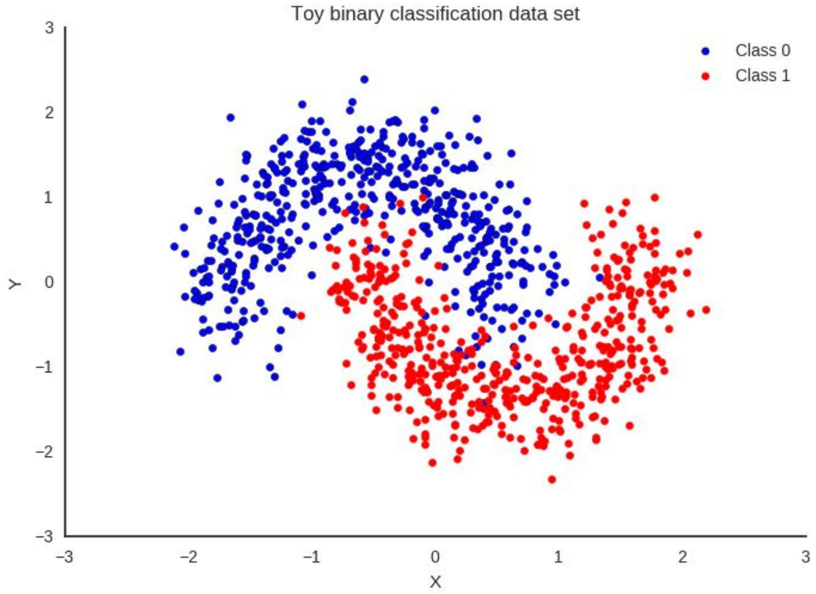


Image source:
https://www.researchgate.net/figure/Example-of-a-binary-classification-for-a-dummy-dataset-with-two-features_fig1_353288055, https://www.kdnuggets.com/2017/04/must-know-evaluate-binary-classifier.html

Binary Classification

- Binary classification aims to distinguish between two classes (e.g., cat and dog, spam or not spam)
- Objective: find a decision boundary that best separates the data points from the two classes

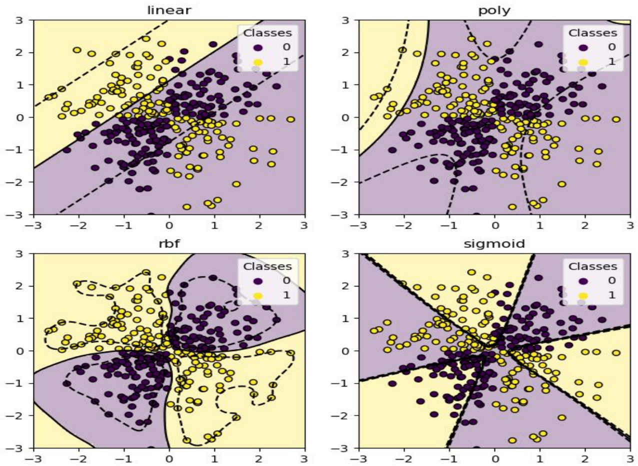


Image source: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_kernels.html

Binary Classification

- Confusion Matrix

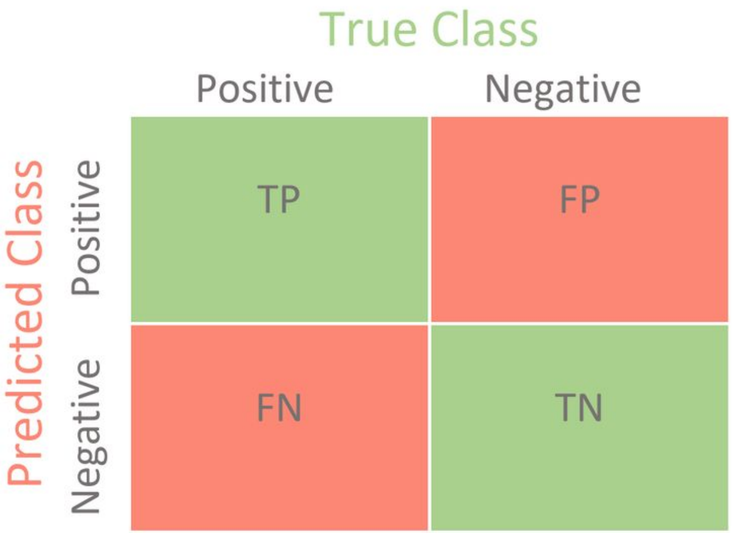


Image source:
https://www.researchgate.net/figure/Confusion-Matrix-for-Binary-Classification-7_fig1_350487701

Binary Classification Metrics

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Accuracy: correct samples / all samples
- Precision: measures how “precise” the model is with predicting the positive class
- Recall: measures how many truly positive samples a model can recall
- If either precision or recall is low, F1 score is low as well
- F1 score demands balance between precision and recall

Binary Classification Metrics

- ROC AUC score: measures how well the model can rank positive samples ahead of negative ones

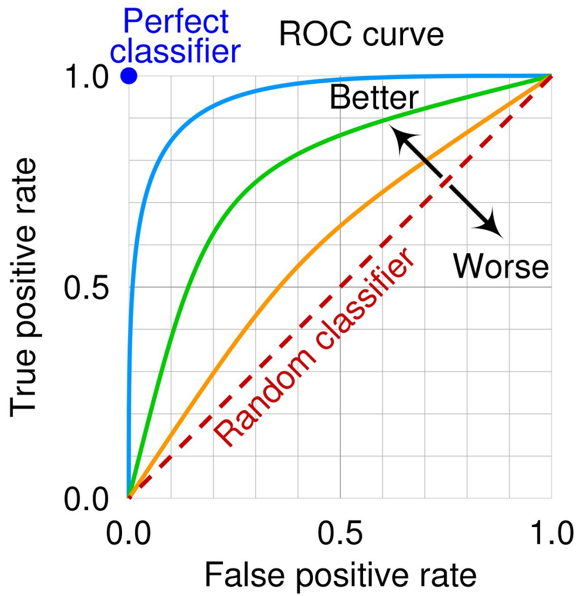


Image source: https://medium.com/@msong507/understanding-the-roc-auc-curve-cc204f0b3441

Binary Classification

- The most simple decision boundary is a straight line
- This is the essence of linear models

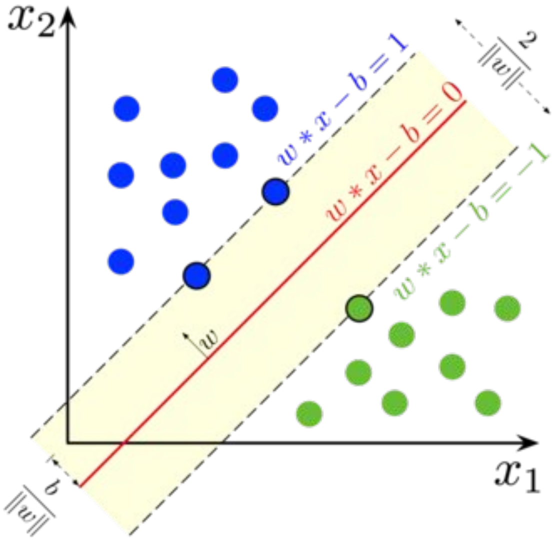
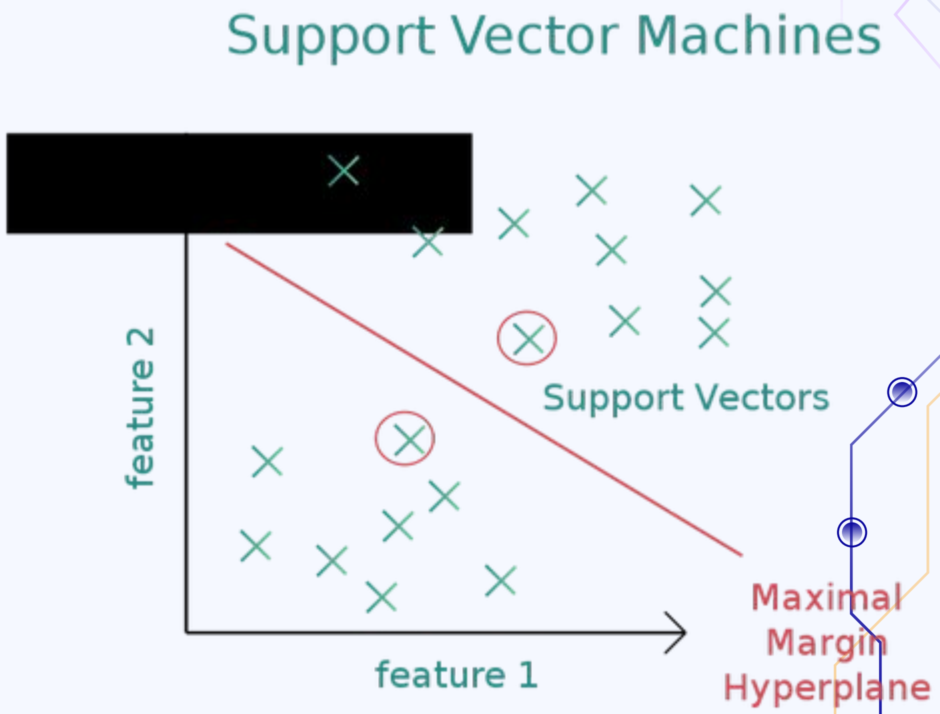


Image source: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:SVM_margin.png

Support Vector Machines

- The <u>support vector machine (SVM)</u> is a linear model training paradigm in which the results are easy to visualize
- Especially if the dataset is <u>linearly separable</u> (i.e., there exists a straight line decision boundary that can perfectly separate all data points of the two classes)



Support Vector Machine

- For linearly separable datasets, the SVM finds the decision boundary that classifies all points correctly AND has the maximum margin
- This means maximizing the distance between the CLOSEST data points to the decision boundary to the boundary itself
- Find the widest “gap” between the two classes

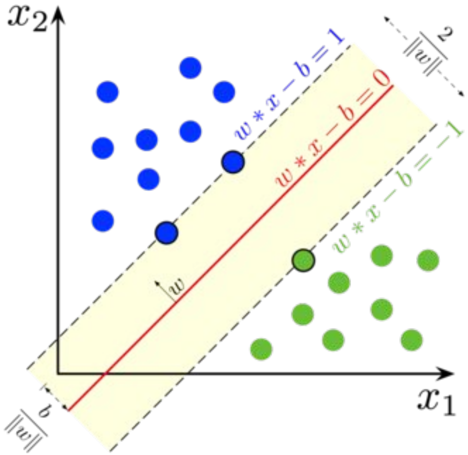


Image source: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:SVM_margin.png

Support Vector Machine

- For datasets that are NOT linearly separable, the SVM needs to balance the classification accuracy and the width of the margin.
- This is controlled by the hyperparameter C
- High C: prioritize correctness of classification
- Low C: prioritize greater margin over classification accuracy

Support Vector Machines

- Train an SVM model and experience the effect of the C hyperparameter here in part 1 of this notebook:
- https://www.kaggle.com/code/lailaicoding/binary-classification

Support Vector Machines

- For datasets that are not linearly separable, another method is to use the “kernel trick”
- Use a “kernel” to map the data to a higher-dimensional space where it may become linearly separable

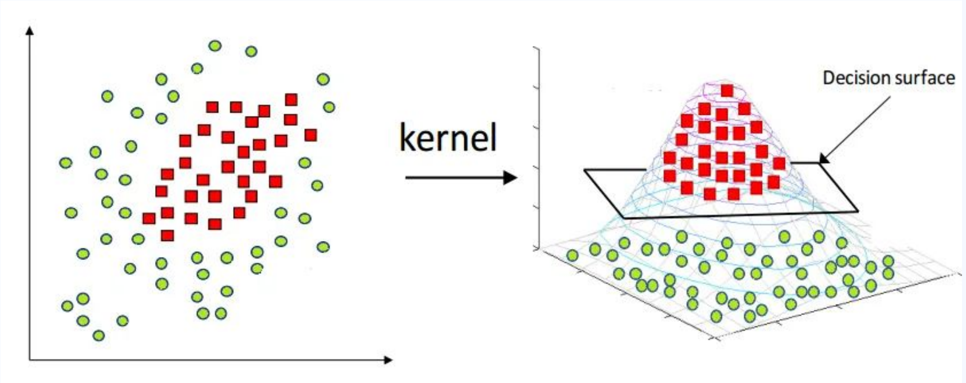


Image source: https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d

Support Vector Machine

- Different taxonomies of decision boundaries allowed when different kernels are used

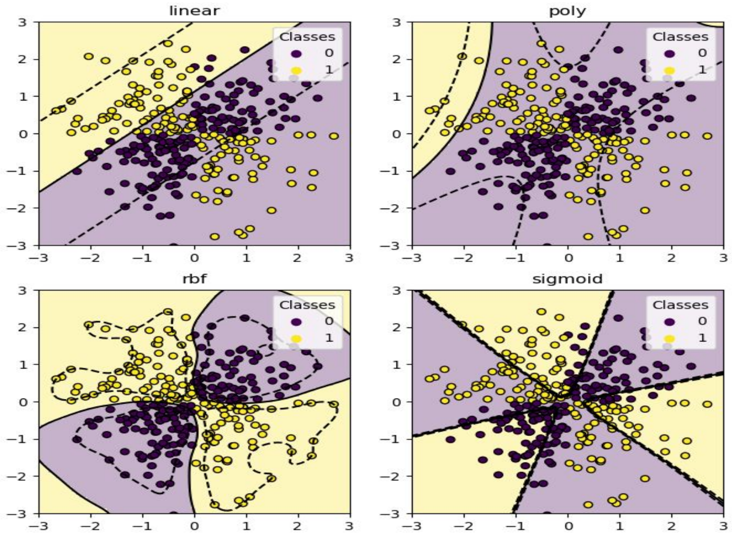


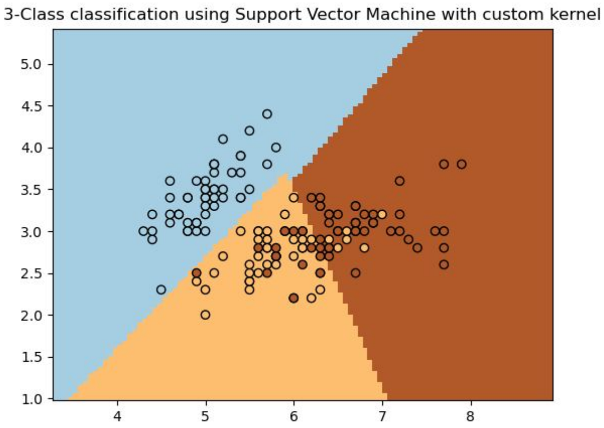
Image source: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_kernels.html

Support Vector Machines

- Experiment with different kernel choices in part 2 of this notebook: https://www.kaggle.com/code/lailaicoding/binary-classification

Multiclass Classification

- Instead of only distinguishing between two classes, distinguish between multiple classes at the same time



- Produce confidence scores for multiple classes at the same time

Image source: https://scikit-learn.org/stable/auto_examples/svm/plot_custom_kernel.html#sphx-glr-auto-examples-svm-plot-custom-kernel-py

Multiclass Classification

- The confidence score can be any real number (positive or negative)
- Convert the scores into a probability distribution using the softmax function
- In the probability distribution, all numbers are positive and they sum to 1
- Take the class with the highest confidence score as the final prediction

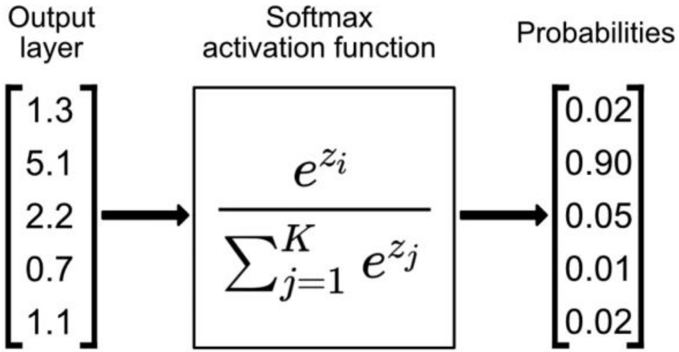
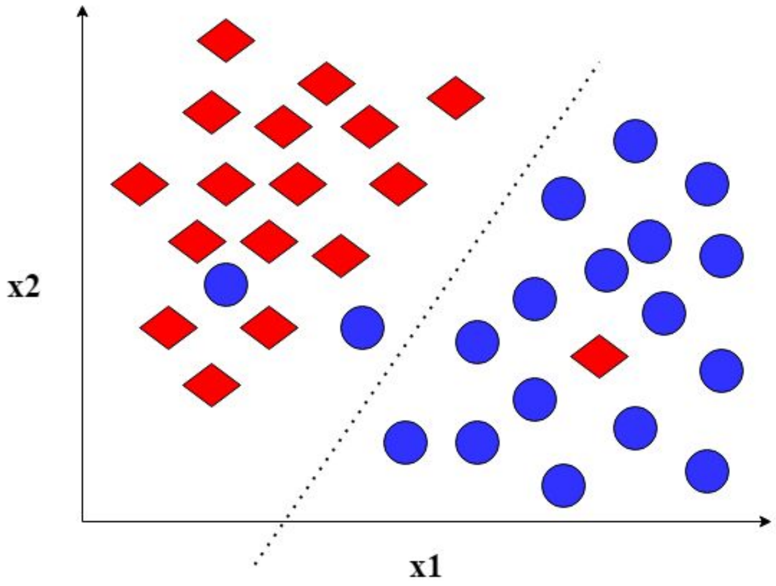


Image source: https://www.singlestore.com/blog/a-guide-to-softmax-activation-function/

Multiclass Classification

- When the confidence score in terms of probability is 90%, we usually can't say that there's a 90% chance that the data point belongs to that class
- The "probabilities" are used to calculate losses, NOT actual probabilities
- We need to calibrate the model to make use of the probabilities (as probabilities) when writing reports

Binary Classification



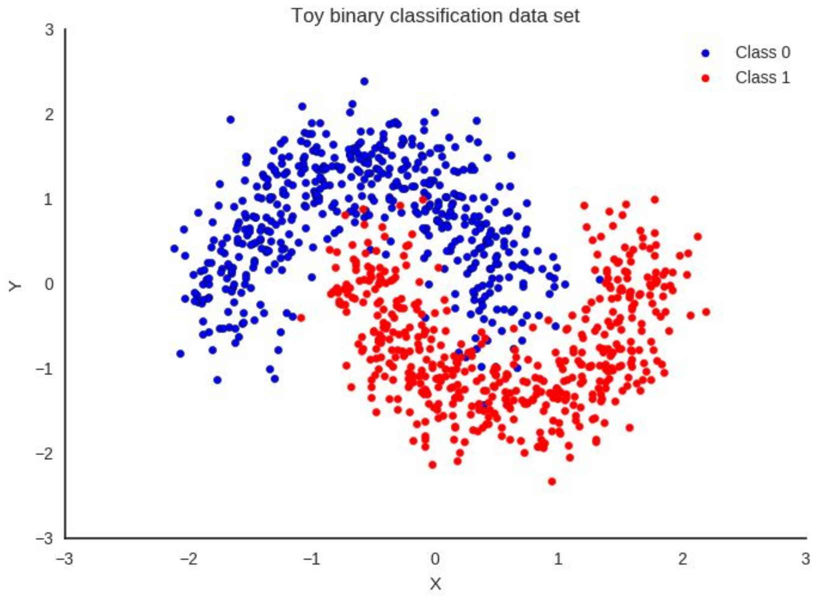


Image source:
https://www.researchgate.net/figure/Example-of-a-binary-classification-for-a-dummy-dataset-with-two-features_fig1_353288055, https://www.kdnuggets.com/2017/04/must-know-evaluate-binary-classifier.html

Binary Classification

- Binary classification aims to distinguish between two classes (e.g., cat and dog, spam or not spam)
- Objective: find a <u>decision boundary</u> that best separates the data points from the two classes

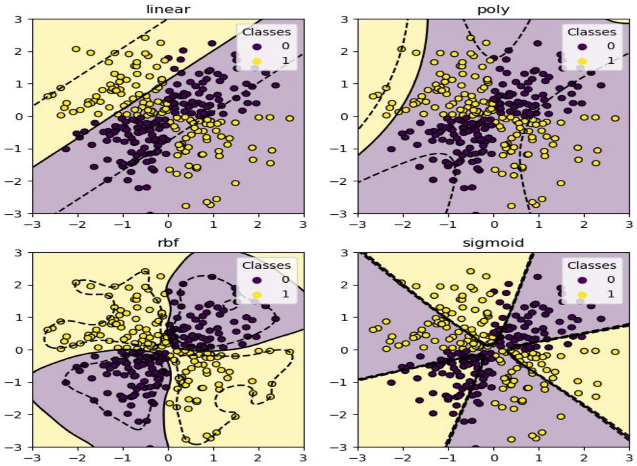


Image source: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_kernels.html

Binary Classification Metrics

- Confusion matrix

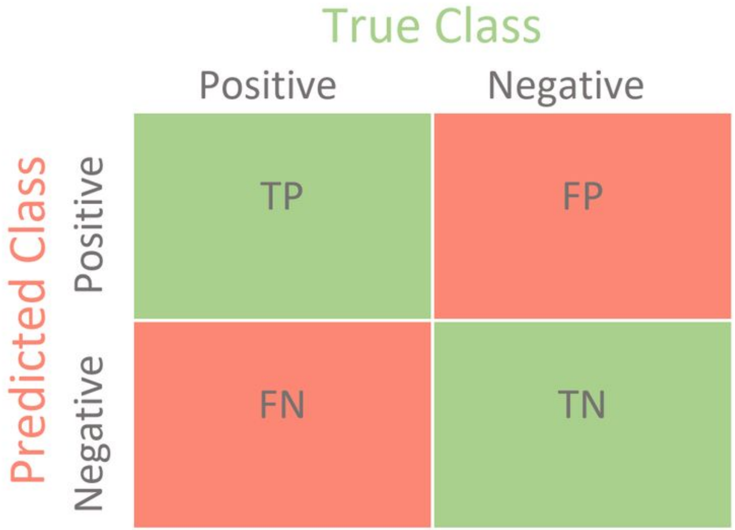


Image source:
https://www.researchgate.net/figure/Confusion-Matrix-for-Binary-Classification-7_fig1_350487701

Binary Classification Metrics

- Accuracy: correct samples / all samples; (TP + TN) / (TP + FP + TN + FN)
- Precision: TP / (TP + FP); measures how “precise” the model is with predicting the positive class
- Recall: TP / (TP + FN); measures how many truly positive samples a model can recall
- F1 score: 2 / (1 / precision + 1 / recall); demands a balance between precision and recall, if either one of them is low then the F1 score will also be low
- All 4 of these metrics output a real number from 0 to 1

Binary Classification

- The most simple decision boundary is a straight line
- This is the essence of linear models

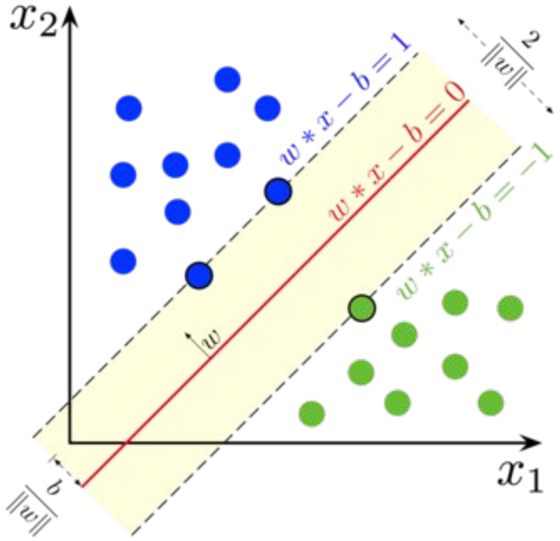


Image source: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:SVM_margin.png

Support Vector Machines

- The support vector machine (SVM) is a linear model training paradigm in which the results are easy to visualize
- Especially if the dataset is linearly separable (i.e., there exists a straight line decision boundary that can perfectly separate all data points of the two classes)

Support Vector Machines

- For linearly separable datasets, the SVM finds the decision boundary that classifies all points correctly AND has the maximum margin
- This means maximizing the distance between the CLOSEST data points to the decision boundary to the boundary itself
- Find the widest “gap” between the two classes

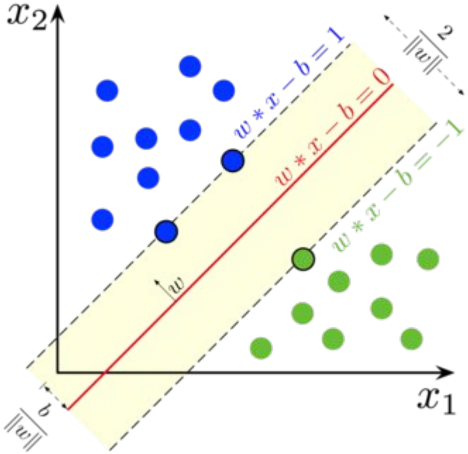


Image source: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:SVM_margin.png

Support Vector Machines

- For datasets that are NOT linearly separable, the SVM needs to balance the classification accuracy and the width of the margin.
- This is controlled by the hyperparameter C
- High C: prioritize correctness of classification
- Low C: prioritize greater margin over classification accuracy

Support Vector Machines

- Train an SVM model and experience the effect of the C hyperparameter here in part 1 of this notebook: https://www.kaggle.com/code/lailaicoding/binary-classification

Support Vector Machines

- For datasets that are not linearly separable, another method is to use the “kernel trick”
- Use a “kernel” to map the data to a higher-dimensional space where it may become linearly separable

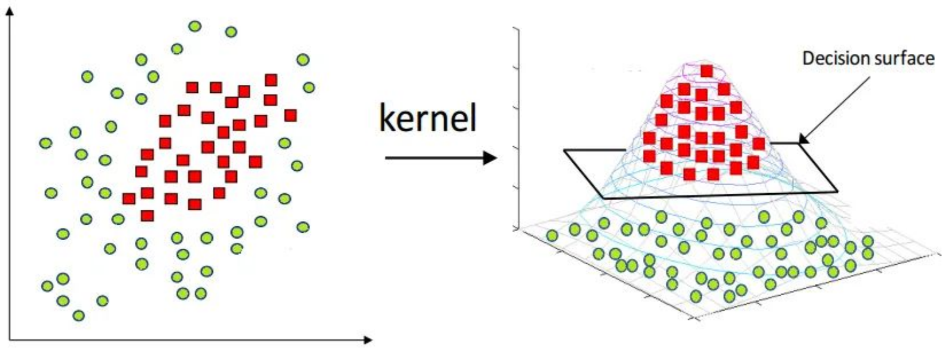


Image source: https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d

Support Vector Machines

- Different taxonomies of decision boundaries allowed when different kernels are used

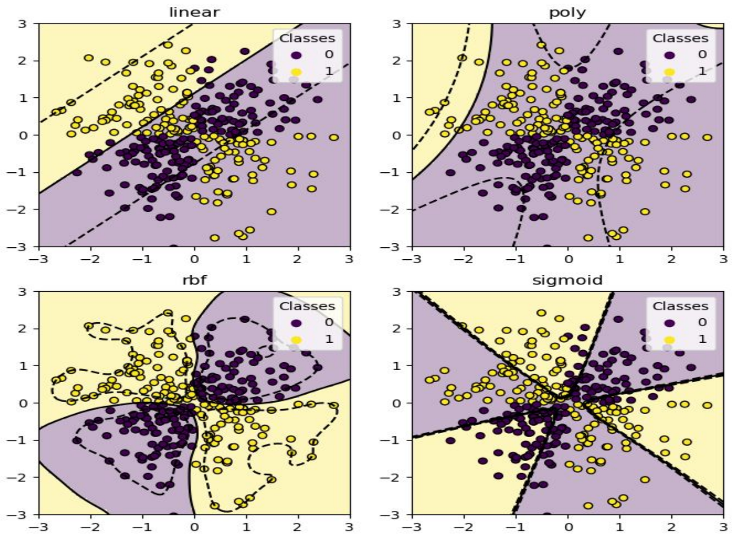


Image source: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_kernels.html

Support Vector Machines

- Experiment with different kernel choices in this notebook: https://colab.research.google.com/drive/1oF2rmTf1ZpuYpI-iuDe3WINEwm6_3x7Q#scrollTo=073a2Jxge5eo

Multi-class Classification

- Instead of only distinguishing between two classes, distinguish between multiple classes at the same time

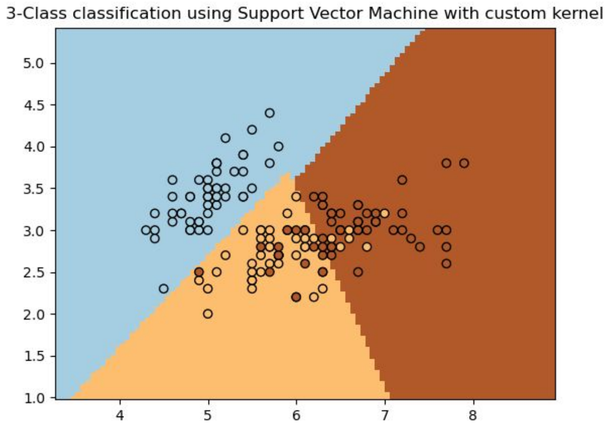


Image source:
https://scikit-learn.org/stable/auto_examples/svm/plot_custom_kernel.html#sphx-glr-auto-examples-svm-plot-custom-kernel-py

Multi-class Classification

- Produce confidence scores for multiple classes at the same time

Multi-class Classification

- The confidence score can be any real number (positive or negative)
- Convert the scores into a probability distribution using the softmax function
- In the probability distribution, all numbers are positive and they sum to 1

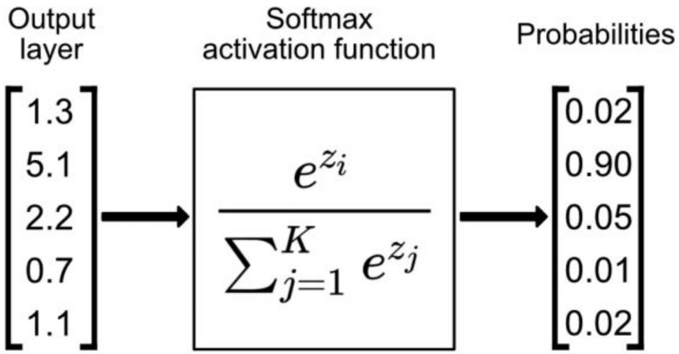


Image source: https://www.singlestore.com/blog/a-guide-to-softmax-activation-function/

Multi-class Classification

- When the confidence score in terms of probability is 90%, we usually can't say that there's a 90% chance that the data point belongs to that class
- The "probabilities" are used to calculate losses, NOT actual probabilities
- We need to calibrate the model to make use of the probabilities (as probabilities) when writing reports