REGRESSION

Intended learning outcomes

- Understand the difference between training data and test data.
- Explain bias-variance tradeoff.
- Explain what is underfitting and overfitting, how to identify them, and how they determine model selection.
- Give different ways to solve overfitting and underfitting.

INTENDED LEARNING OUTCOMES

Optimization

- Give examples of loss functions, and define empirical risk in terms of the loss function.
- List two general types of algorithms used in optimization, e.g. exact solution, and gradient descent. Outline the broad steps involved in each of them.
- Explain why framing a problem as convex optimization is highly desirable, in terms of speed and local minima.
- Explain the motivation behind performing stochastic gradient descent, rather than traditional gradient descent.

INTENDED LEARNING OUTCOMES

Multivariate Linear Regression

- State the model and the training loss.
- Explain how the ‘constant feature’ trick can be used to reduce the problem to one without the constant parameter θ_0.
- Describe two training algorithms that may be applied.
- Derive the gradient of the training loss.
- Derive the formula for the exact solution.
- Describe two potential weaknesses of the exact solution, and possible solutions for these weaknesses.
- Apply the above algorithms to a given data set.

INTENDED LEARNING OUTCOMES

Regularization

- Explain why regularization can help with generalization.
- State the training loss and test loss in ridge regression.
- Identify the regularizer and regularization parameter in the training loss of a given machine learning problem.
- Explain why regularization solves the invertibility problem in traditional linear regression.
- Describe the difference in gradient descent between traditional and regularized linear regression.
- Describe how the test loss and training loss varies with the regularization parameter.

CLASSIFICATION

Intended learning outcomes

- Understand what decision boundaries and regions are with respect to a classifier.
- Define what it means for a dataset to be linearly separable, and give an example of a dataset that is not.

INTENDED LEARNING OUTCOMES

Logistic Regression

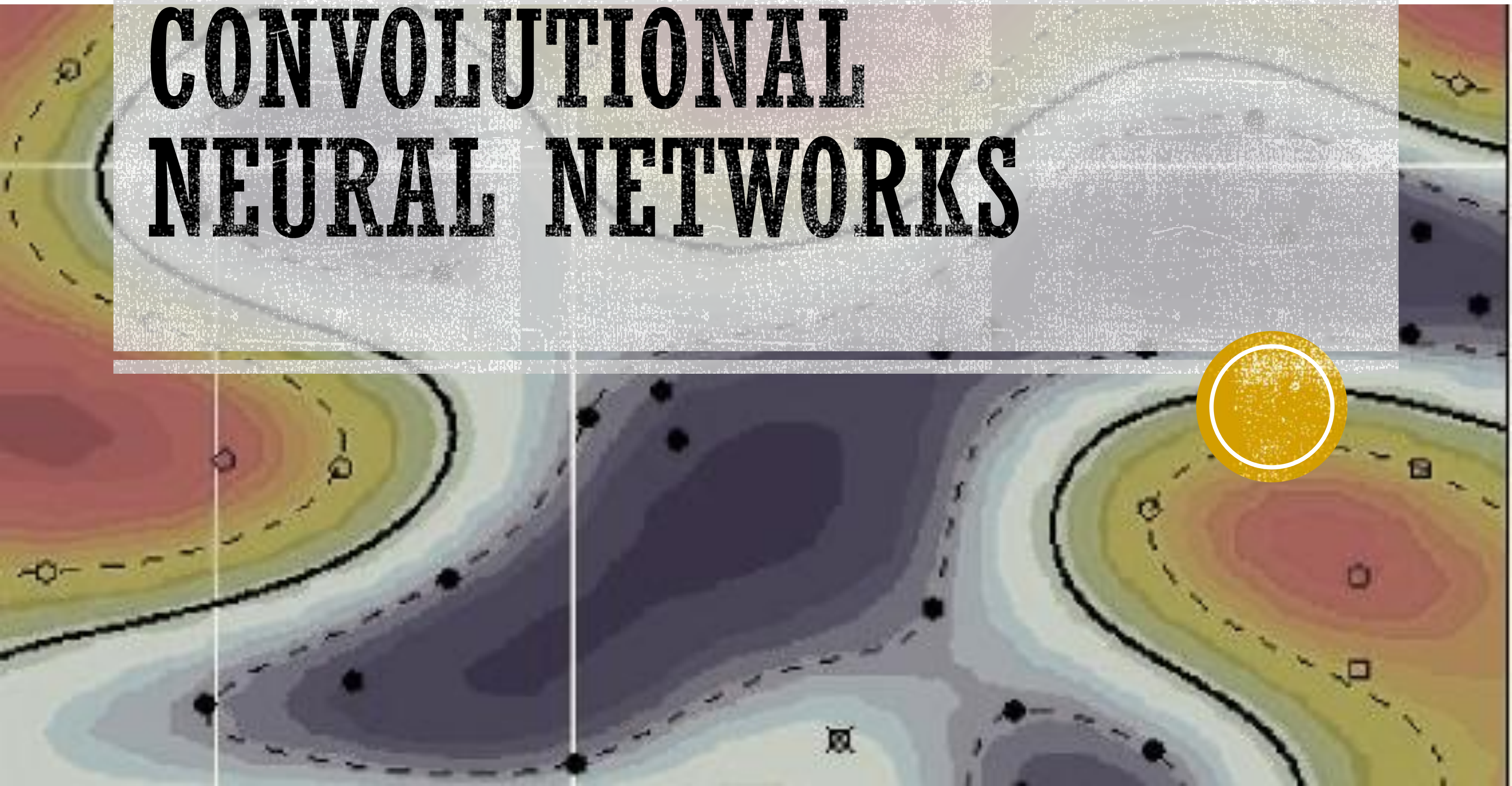
- Write down the sigmoid function and logistic loss function.
- Describe the model as a set of sigmoid neurons which predict the probability of the labels given the features.
- Derive the label predictions from the label probabilities. Describe the decision boundary.
- Derive the training loss from the likelihood of the data, and write it in terms of the logistic loss.
- Derive the training gradient, and describe an algorithm for performing logistic regression.

DEEP LEARNING

Intended learning outcomes

- Describe how the output of an artificial neuron relates to its inputs.
- Give examples of commonly used activation functions.
- Know how to efficiently compute the error signal in backpropagation, and be able to explain why it is done this way.
- Understand the statement of the universal approximation theorem.
- Give reasons for the recent successes of deep learning, and explain how it differs from classical machine learning.

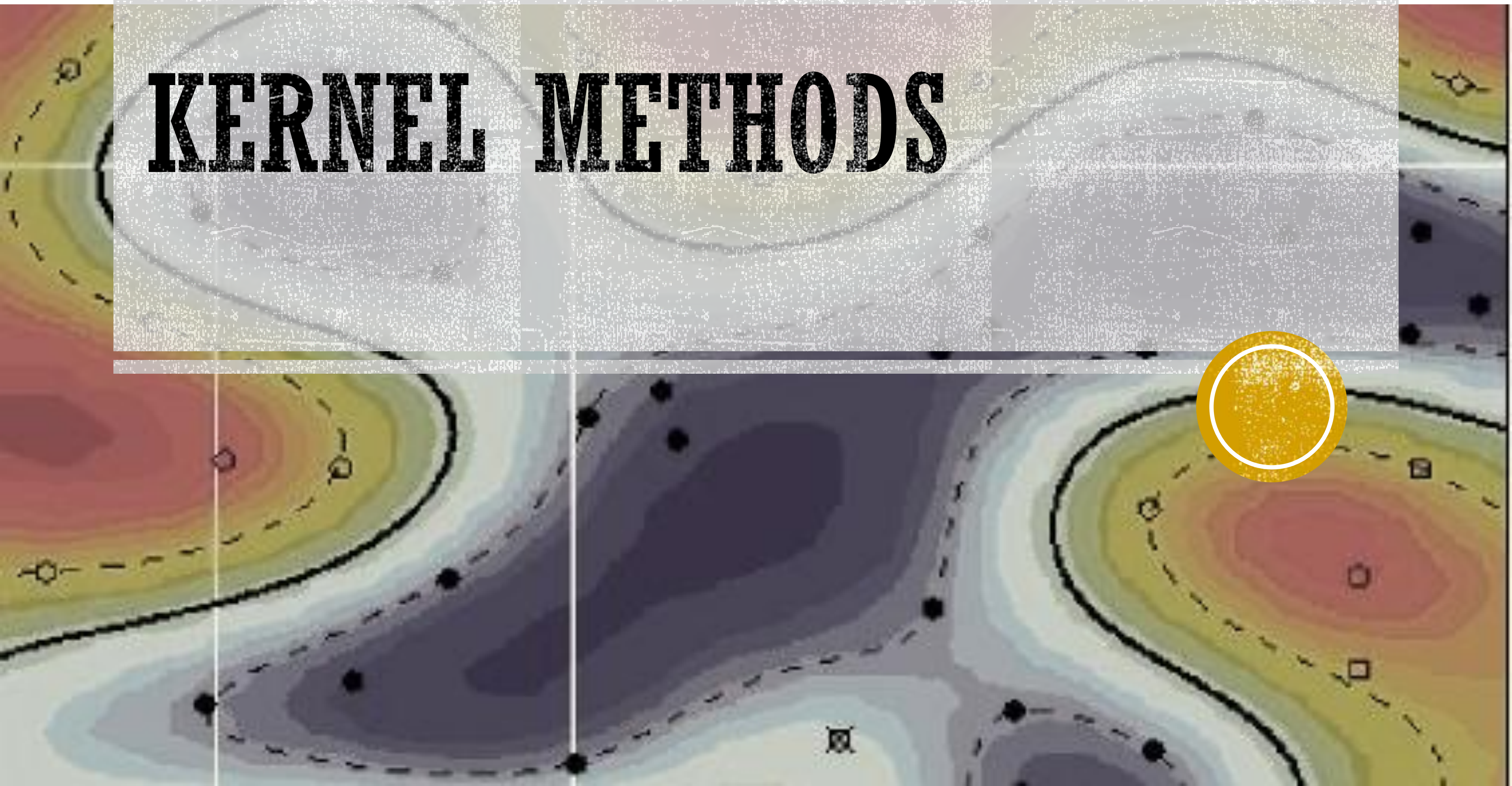
CONVOLUTIONAL NEURAL NETWORKS



Intended learning outcomes

- Understand how convolution is performed on 2D arrays.
- Explain why convolution is used for image recognition, and describe how it builds up a hierarchy of features.
- Know the definitions of the terms "stride", "padding" and "filter size".

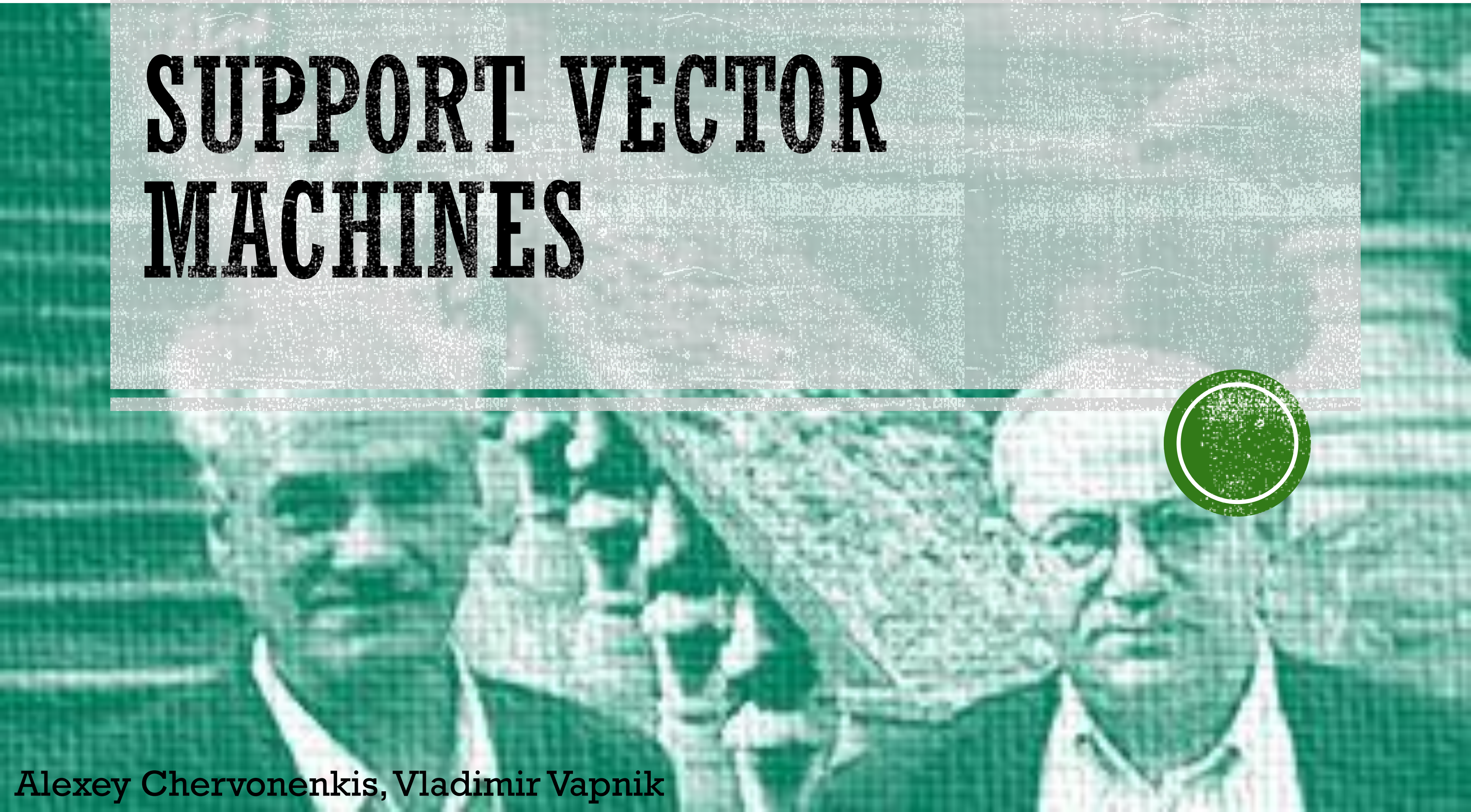
KERNEL METHODS



Intended learning outcomes

- Give the definition of a kernel function, and be able to ascertain what a valid kernel function is.
- Describe how feature maps change a linear algorithm into a non-linear one.
- Explain what the kernel trick is, and how it bypasses having to define a computationally intensive feature map.
- Understand how kernels typically arise in dual representations of the problem.

SUPPORT VECTOR MACHINES



Alexey Chervonenkis, Vladimir Vapnik

INTENDED LEARNING OUTCOMES

Support Vector Machines

- Write down the primal problem, and explain how it is derived from the maximum margin problem.
- Write down the dual problem, and identify the kernel. Describe how the optimal θ is derived from the $\alpha_{x,y}$'s. Describe in terms of the $\alpha_{x,y}$'s, how to do prediction.
- Define support vectors, both geometrically and in terms of the $\alpha_{x,y}$'s. Recognize that most of the $\alpha_{x,y}$'s are zero.

INTENDED LEARNING OUTCOMES

Extensions

- Describe the dual problem for the SVM with offset.
- Describe the primal problem for SVM with slack variables. Show that the primal is equivalent to regularized hinge loss. Explain how the regularizing parameter λ affects the margins. Describe the dual problem in terms of box constraints.

Gaussian processes for regression and optimization

Intended learning outcomes

- Define a Gaussian process and understand them as infinite-dimensional generalizations of multivariate Gaussian random variables.
- Understand the difference between parametric and non-parametric methods.
- Use the formulas for conditional Gaussian distributions to perform Gaussian process prediction.
- Define what an acquisition function is and list commonly used examples.
- Understand how acquisition functions are used in conjunction with Gaussian process prediction to do Bayesian optimization.

Graphical models

Intended learning outcomes

- Understand conditional independence of random variables.
- Understand what a Markov random field is, and the properties its graph and joint distribution must satisfy.
- Describe what d-separation is.
- Describe a Bayesian network, and the properties it must fulfill.