Quiz 1

Graded

Student

Alaa Alghwiri

View or edit group

Total Points

6 / 6 pts

Question 1

Probability of Classes — **1** / 1 pt

✔ **+ 1 pt** Correct

Question 2

Basis Functions — **1** / 1 pt

✔ **+ 1 pt** Correct

Question 3

Function Approximation — **1** / 1 pt

✔ **+ 1 pt** Correct

Question 4

Universal Function Approximation — **1** / 1 pt

✔ **+ 1 pt** Correct

Question 5

Binary Classification — **1** / 1 pt

✔ **+ 1 pt** Correct

Question 6

Gradienet Descent and Linear Regression — **1** / 1 pt

✔ **+ 1 pt** Correct

Q1 Probability of Classes

1 Point

We proposed the objective functions $l(w) = -\prod_{i=1}^{n} \Pr(\hat{Y} = y_i | X = x_i)$ and $l'(w) = -\sum_{i=1}^{m} \ln \Pr(\hat{Y} = y_i | X = x_i)$ for binary classification. What is true about these two objective functions?

- [x] They both of the same optima
- [ ] If these functions are optimized with gradient descent, they will both find a solution in the same number of iterations (number of steps of gradient descent).
- [x] The log probability version, $l'(w)$, has a better gradient when misclassifications occur.
- [ ] The gradients do not point in the same direction
- [x] The gradient of the first one, $\nabla l(w)$, goes towards zero when there are large misclassifications.

Q2 Basis Functions

1 Point

If we create an approximation $f(x, w) = \phi(x)^\top w$ of f_*, where $\phi \colon \mathbb{R}^n \to \mathbb{R}^m$ is a basis function (e.g., polynomial basis, Fourier basis, or binning), and $w \in \mathbb{R}^m$, we say that this is linear function approximation because:

- ( ) ϕ is a linear function
- ( ) we do not call this linear function approximation because ϕ is a nonlinear function
- (●) the function is linear in the parameters w
- ( ) the function can only represent linear transformations of x

Q3 Function Approximation
1 Point

What are the reasons for approximating a function? Mark all that apply

- [ ] Copy the target function exactly
- [x] Be able to imitate the behavior of some complex process like human decision-making
- [x] Create a faster version of the original function
- [x] Create a known function that we can use when the target function is unknown

Q4 Universal Function Approximation
1 Point

With an appropriate Basis function (or neural network), the universal function approximation theorem says

- (●) with a possibly infinite number of basis functions, any function can be represented to an arbitrarily small precision.
- ( ) we can optimize the function or neural network using gradient descent to find a perfect fit
- ( ) that with enough basis functions, we can approximate any function to an arbitrarily small precision
- ( ) if we have infinite data, we can approximate any function to an arbitrarily small precision

Q5 Binary Classification

1 Point

In binary classification, we want to maximize the number of correct classifications or, equivalently, minimize the number of misclassifications. We proposed the objective $l(w, b) = -\sum_{i=1}^{m} y_i \text{sign}(w^\top x + b)$, which counts up the number of correct classifications and subtracts the number of incorrect classifications. Which of the following are reasons we proposed a different objective that aimed to maximize the probability of correct classifications?

- [ ] It is not possible to find the optimal parameters w and b.
- [x] The function is not differentiable.
- [ ] The minimum of this function does not correspond to achieving perfect classification.
- [ ] Maximizing probabilities finds a better solution because the classes are more likely.

Q6 Gradienet Descent and Linear Regression

1 Point

In linear regression with the objective function $l(w) = \frac{1}{2}\mathbf{E}\left[(f(X, w) - Y)^2\right]$, we can find the optimal weights with the expression $w_* = \mathbf{E}\left[\phi(x)\phi(x)^\top\right]^{-1}\mathbf{E}[\phi(X)Y]$. If we perform gradient descent on this objective function, we are not guaranteed to find w_* because gradient descent only finds a local optimum and not a global optimum.

- ○ true
- ● false