CS209: Machine Learning - Question Bank

Based on the provided syllabus, here's a comprehensive question bank organized by topics with theory, derivation, and numerical problems.

TOPIC 1: Machine Learning Concepts

Theory Questions:

1. **T1.1** Define machine learning and explain the difference between artificial intelligence, machine learning, and deep learning with examples.
2. **T1.2** What are the key components of designing a learning system? Explain each component with suitable examples.
3. **T1.3** Compare and contrast supervised, unsupervised, and semi-supervised learning with real-world applications.
4. **T1.4** Explain the concept of inductive bias in machine learning. Why is it necessary?
5. **T1.5** What is the No Free Lunch theorem? What are its implications for machine learning?
6. **T1.6** Discuss the bias-variance tradeoff. How does it relate to overfitting and underfitting?
7. **T1.7** Explain cross-validation techniques. Compare k-fold, leave-one-out, and stratified cross-validation.
8. **T1.8** What is the curse of dimensionality? How does it affect different machine learning algorithms?
9. **T1.9** Explain the difference between parametric and non-parametric learning algorithms with examples.
10. **T1.10** What are the assumptions made in linear regression? When might these assumptions be violated?

Derivation Questions:

1. **D1.1** Derive the normal equation for linear regression starting from the least squares cost function: $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
2. **D1.2** Derive the gradient descent update rule for linear regression. Show the partial derivatives with respect to each parameter.
3. **D1.3** Derive the cost function for logistic regression using maximum likelihood estimation.
4. **D1.4** Derive the sigmoid function and show that its derivative is σ(z)(1-σ(z)).
5. **D1.5** Prove that the logistic regression cost function is convex.

Numerical Problems:

1. **N1.1** Given the dataset:

x	y
1	2

x	y
2	4
3	6
4	8

Calculate the parameters θ_0 and θ_1 using the normal equation.

2. **N1.2** For logistic regression with θ = [0.5, -1.2, 0.8] and input x = [1, 2, 3]:
   - Calculate the predicted probability $P(y=1|x)$
   - If actual y=1, calculate the cost
3. **N1.3** Perform one iteration of gradient descent for linear regression with:
   - Initial θ = [0, 0]
   - Learning rate $\alpha = 0.01$
   - Training data: (1,1), (2,3)
4. **N1.4** Calculate the R^2 score for predictions $\hat{y}$ = [2.1, 3.9, 6.2] and actual y = [2, 4, 6].
5. **N1.5** Given training error = 0.15 and validation error = 0.25, determine if the model is overfitting, underfitting, or well-fitted. Suggest remedies.

TOPIC 2: Decision Theory & Information Theory

Theory Questions:

1. **T2.1** Explain Bayes' theorem and its significance in machine learning. Provide a real-world example.
2. **T2.2** What is entropy in information theory? How is it related to uncertainty?
3. **T2.3** Define mutual information and explain its role in feature selection.
4. **T2.4** Explain the concept of KL-divergence and its properties.
5. **T2.5** What is the difference between frequentist and Bayesian approaches to probability?
6. **T2.6** Explain the maximum entropy principle and its applications.
7. **T2.7** What are conjugate priors? Why are they useful in Bayesian inference?
8. **T2.8** Explain the concept of sufficient statistics with examples.
9. **T2.9** What is the central limit theorem and how is it relevant to machine learning?
10. **T2.10** Discuss different probability distributions commonly used in machine learning.

Derivation Questions:

1. **D2.1** Derive the formula for entropy: $H(X) = -\sum p(x)\log_2 p(x)$
2. **D2.2** Prove that entropy is maximized when all outcomes are equally likely.
3. **D2.3** Derive the formula for conditional entropy: $H(Y|X) = \sum p(x)H(Y|X=x)$
4. **D2.4** Show that mutual information $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
5. **D2.5** Derive the KL-divergence formula and show that it's always non-negative.

Numerical Problems:

1. **N2.1** Calculate the entropy of a coin with P(Heads) = 0.7.
2. **N2.2** Given the joint probability distribution:

X\Y	0	1
0	0.3	0.2
1	0.1	0.4

   Calculate $H(X)$, $H(Y)$, $H(X,Y)$, $H(X|Y)$, and $I(X;Y)$.
3. **N2.3** Calculate the KL-divergence between distributions P = [0.5, 0.3, 0.2] and Q = [0.4, 0.4, 0.2].
4. **N2.4** For a dataset with classes: [A, A, B, B, B, C], calculate the entropy.
5. **N2.5** Given P(Disease) = 0.01, P(Test+|Disease) = 0.95, P(Test+|No Disease) = 0.05, calculate P(Disease|Test+).

TOPIC 3: Bayesian Learning

Theory Questions:

1. **T3.1** Explain the Bayesian approach to learning. What are prior, likelihood, and posterior?
2. **T3.2** What is the difference between MAP (Maximum A Posteriori) and MLE (Maximum Likelihood Estimation)?
3. **T3.3** Explain the concept of conjugate priors with examples for different likelihood functions.
4. **T3.4** What is Bayesian model selection? How does it differ from frequentist model selection?
5. **T3.5** Explain the concept of hyperparameters in Bayesian learning.
6. **T3.6** What is the evidence (marginal likelihood) in Bayesian inference? Why is it difficult to compute?
7. **T3.7** Explain Bayesian linear regression and how it differs from classical linear regression.
8. **T3.8** What is the Bayesian information criterion (BIC)? How is it derived?
9. **T3.9** Explain the concept of empirical Bayes and when it's used.
10. **T3.10** What are the advantages and disadvantages of the Bayesian approach?

Derivation Questions:

1. **D3.1** Derive the posterior distribution for Bayesian linear regression with Gaussian prior and likelihood.
2. **D3.2** Show that the Beta distribution is conjugate to the Binomial likelihood.
3. **D3.3** Derive the MAP estimate for a parameter with Gaussian prior and Gaussian likelihood.
4. **D3.4** Prove that MLE is a special case of MAP with uniform prior.
5. **D3.5** Derive the predictive distribution in Bayesian linear regression.

Numerical Problems:

1. **N3.1** Given a Beta(2,3) prior for coin bias and observing 5 heads in 8 tosses, find the posterior distribution.
2. **N3.2** For Bayesian linear