# Tools and Techniques for Machine Learning
# Homework 3: Thompson sampling and counterfactual policy evaluation

**Instructions**: Your answers to the questions below, including plots and mathematical work, should be submitted as a single PDF file. It's preferred that you write your answers using software that typesets mathematics (e.g. LaTeX, LyX, or Jupyter), though if you need to you may scan handwritten work. For submission, you can also export your Jupyter notebook and merge that PDF with your PDF for the written solutions into one file. **Don't forget to complete the Jupyter notebook as well, for the programming part of this assignment**.

## 1 Derviation of importance-weighted reward imputation

Suppose we have a contextual bandit where context $X \in \mathcal{X}$ has probability density function $p(x)$ and reward vector $R \in \mathbb{R}^k$ has conditional distribution given by $P_{R|X}$. We want to use the direct method to evaluate the performance of a static policy $\pi$. That is, we want to use

$$\begin{aligned}\hat{V}_{\mathrm{dm}}(\pi) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{k}\hat{r}(X_i,a)\pi(a \mid X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{A_i\sim\pi(\cdot|X_i)}\left[\hat{r}(X_i,A_i)\right],\end{aligned}$$

where $\hat{r}(x,a)$ is some estimate for $\mathbb{E}\left[R(A) \mid X=x, A=a\right] = \mathbb{E}\left[R(a) \mid X=x\right]$ and

$$(X_1, A_1, R_1(A_1)), \ldots, (X_n, A_n, R_n(A_n))$$

is the logged bandit feedback from static policy $\pi_0$ on the same contextual bandit distribution. The "naive" approach to fitting $\hat{r}$ from some hypothesis space $\mathcal{H}$ is least squares:

$$\hat{r} = \arg\min_{r\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\left(r(X_i,A_i) - R_i(A_i)\right)^2.$$

1. With this approach, what is the covariate distribution in training? Explain why we have a covariate shift between the train and target distribution.
2. Give an importance-weighted objective function $J(r)$ for finding $\hat{r}$, and use the change of measure theorem to show that $\mathbb{E}\left[J(r)\right] = \mathbb{E}\left[r(X,A) - R(A)\right]^2$, where $X \sim p(x)$, $R \mid X \sim P_{R|X}$ and $A \mid X \sim \pi(a \mid x)$. In other words, the objective function is an unbiased estimate of the expected square loss (i.e. the risk) of $r$ w.r.t. the target distribution.

## 2 Optimizing 0/1 loss for binary classification

1. Suppose we're trying to predict a binary event with outcome space $\mathcal{Y} = \{0, 1\}$. The action space is also $\mathcal{A} = \{0, 1\}$, but we make randomized predictions with $\mathbb{P}(A = 1) = \pi$. Suppose we know that $\mathbb{P}(Y = 1) = p$. We want to find $\pi \in [0, 1]$ that minimizes our expected loss $\mathbb{E}\ell(A, Y)$. What is the optimal $\pi$ for the 0/1 loss: $\ell(a, y) = \mathbb{1}[a \neq y]$?

2. Now consider the corresponding probabilistic prediction problem, where the action space is $\mathcal{A} = [0, 1]$, which is supposed to be a prediction for $\mathbb{P}(Y = 1)$. Give a loss function $\ell(a, y)$ for this action space and outcome space $\mathcal{Y} = \{0, 1\}$ such that the expected loss is equivalent to the expected loss of the previous problem. That is, find $\ell(a, y)$ such that $\mathbb{E}_{Y\sim\text{Ber}(p)}\ell(\pi, Y) = \mathbb{E}_{A\sim\text{Ber}(\pi),Y\sim\text{Ber}(p)}\mathbb{1}[A \neq Y]$. (Hint: Here and below, it may be helpful to note that if

$$f(y) = \begin{cases} a & \text{when } y = 0 \\ b & \text{when } y = 1, \end{cases}$$

then $f(y) = a^{(1-y)}b^y$.)

3. Consider the conditional probability modeling setting with input space $\mathcal{X} = \mathbb{R}^d$ and outcome space $\mathcal{Y} = \{0, 1\}$. We want to predict the probability of the outcome 1 for any input $x \in \mathcal{X}$. The logistic regression model is $\mathbb{P}(Y = 1 \mid X = x; w) = \phi(w^T x)$, where $\phi(\eta) = 1/(1 + e^{-\eta})$ (the standard logistic function). We typically fit $w \in \mathbb{R}^d$ by minimizing the negative log-likelihood of $w$ for some data $\mathcal{D} = ((X_i, Y_i))_{i=1}^n$ samplied i.i.d. from the data generating distribution $P$. The negative log likelihood objective is

$$J_{\text{nll}}(w) \quad = \quad -\left[\sum_{i=1}^{n} Y_i \log \phi(w^T X_i) + (1 - Y_i) \log\left(1 - \phi(w^T X_i)\right)\right].$$

Now consider the setting of supervised learning with a 0/1 loss function: $\ell(a, y) = \mathbb{1}[a \neq y]$. Give an expression for the expected loss of a randomly sampled $(X, Y) \sim P$ when the action $A \in \{0, 1\}$ is drawn randomly from the logistic regression model described above, namely $\mathbb{P}(A = 1 \mid X = x; w) = \phi(w^T x)$

4. Give an objective function in terms of the data $\mathcal{D}$ for finding the $w$ that optimizes the expected loss you gave in the previous problem. The objective function should be an unbiased estimate for $\mathbb{E}_w\ell(A, Y)$. Is this objective function equivalent to the negative log-likelihood objective function described above? If not, how might you expect the logistic regression models resulting from the two different objective functions to compare? Consider the case of large datasets and very expressive feature spaces. (Hint: log loss is a proper scoring rule.)