







Training: 49000 games done.

100 test games: 0% lost, 39% tied, 61% won.

Training: 50000 games done.

100 test games: 0% lost, 36% tied, 64% won.

10000 test games: 1.83% lost, 44.41% tied, 53.76% won.

>>

ב. לאחר נסיונות רבים (מאוד), ניכר כי הרשת שהניבה את התוצאות הטובות ביותר הייתה תחת מדיניות *softmax* ובעלת הפרמטרים שהוגשו. סביר מאוד כי מול יריב דטרמניסטי, עם מספיק אימון הייתה עדיפה מדיניות דטרמניסטית (כמו *greedy* למשל) אשר תאפשר למידה מלאה של היריב (אין אלמנט אקראי) ופיתוח אסטרטגיה למיקסום הגמול מולו.

ג. גם לאחר נסיונות רבים עם פרמטרים שונים, ניכר כי הרשת אינה בעלת ביצועים אופטימליים. על אף חוסר האופטימליות, בחרתי להשאיר בתצורה זו עקב היציבות הגבוהה בתוצאות היחסית טובות ויעילותה הכוללת (מגיע *gap* במבחן הסופי של כ-51%). ניתן לחשוב על מגוון סיבות מדוע הרשת אינה אופטימלית, אתמקד בשתיים העיקריות לטעמי הקשורות באופן מסוים האחת בשנייה: ראשית, מדיניות היריב הינה אקראית לחלוטין במרבית מהלכיו (בכולם פרט לאחרון במשחקים בו יש לו הזדמנות לנצח). ניסיון ללמוד אקראיות מוחלטת הינו בעייתי ונידון לכישלון מסוים מראש. שנית, מאחר ומרבית מהלכיו של היריב אקראיים לחלוטין, נדרש כוח חישובי משמעותי מאוד על מנת לייצר למידה עבור כמעט כל מצבי הלוח האפשריים. ניסיון שכזה היה דורש זמן ומשאבים רבים הרבה יותר אשר נחסכו במטלה זו על מנת לעמוד בדרישות זמני הריצה.